Title
Biophysical Characterization of Ribosome Nascent Chain Energy Landscapes

Author
Samelson, Avi Jacob

Publication Date
2017

Peer reviewed|Thesis/dissertation

Biophysical Characterization of Ribosome Nascent Chain Energy Landscapes

By

Avi Jacob Samelson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susan Marqusee, Chair
Professor Jaime H.D. Cate
Professor David F. Savage
Professor David E. Wemmer

Summer 2017

Abstract

Biophysical Characterization of
Ribosome Nascent Chain Energy Landscapes

By

Avi Jacob Samelson

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Susan Marqusee, Chair

Protein folding is necessary for proper cellular and organismal function. The disruption of accurate protein folding is highly detrimental to numerous cellular processes and can lead to cell death and disease. A major protein homeostasis node in the cell is the ribosome, which interacts with a number of quality control factors that help guarantee protein folding fidelity. For many years, co-translational folding has been investigated mainly within the context of cellular proteostasis and quality control. Unfortunately, how the ribosome and translational machinery modulates the nascent chain itself has, for the most part, remained a mystery. It is essential to understand how translation changes a protein's energy landscape during translation in order to understand how folding fidelity can be guaranteed *in vivo*. Here, I introduce new methods for studying the folding thermodynamics and kinetics of stalled nascent chains and pioneer a new assay for studying folding kinetics during translation in real-time. In addition, I characterize the folding landscape of the protein HaloTag and determine that the process of translation itself increases HaloTag's folding efficiency by modulating its folding trajectory. The methods developed herein, as well as novel observations about the biophysical properties of nascent chains reported here, pave the way for understanding protein folding *in vivo*.

Dedication

To my mother, Judy Lieberman
To Tom Alber
To Howard Schachman

צ״ל

Acknowledgements

First, I have to thank my dad and sister: you have given me so much, I hope I can make you proud. I need to thank my mother's parents, Zade Jerry and Bubbe Libby for teaching me the meaning of true love and the satisfaction of working with your hands, and my dad's parent's Zade Bill and Bubbe Rose for teaching me the importance of persistence, hope, and hard work. I also am lucky enough to have an extremely large extended family, many of whom live in the bay area: uncle Mike and aunt Marlene, my cousins Jon, Trish, Daniel, Emma, Hannah, Noel, Robbie, Ry, Aaron, uncle Jerry and Nancy. I have to also thank the rest of my extended family including Abi, Adam and their new baby Elijah as well as the rest of my parents' siblings, cousins and their children.

I am sure that I would not be here without James Fraser and his wife Rachel Zunder. You both taught me so much about what it means to be a scientist and a friend.

I have to also thank my old friends: Nate, Rina, Emma, Caryn, Ellen to name a few; my college friends, especially Adam, Thomas, Sean, Fizza, Pat, Anand, Zoya, Kate, Peter, Aaron; and all of my classmates here at Berkeley especially John Blair, Andrew, Jeannette, Evan and Ellen.

Here at Berkeley, I have had the pleasure of being able to walk around Stanley Hall and be able to find expertise and equipment. I have had many teachers just wandering around the building trying to find a way to do an experiment. I have to specifically thank Andy Martin, John Kuriyan, Dave Wemmer, Jamie Cate and Jennifer Doudna and the members of their labs especially Stephen Floor, who basically was a fifth member of my thesis committee.

I owe so much to past and present members of the Marqusee Lab. Sabriya Rosemond was the best rotation mommy I could ever have and taught me so much. Diana Koulechova, Laura Rosen and Katies Hart and Tripp were big sisters to me. Sam Leachman made lab both fun and funny. Madeleine Jensen and Emily Guinn taught me so much and were great ribosome buddies. Shion Lim and Eric Bolin affirmed the joy of collaborative science for me and were always willing to talk protein folding kinetics through with me. Helen Hobbs, Charlotte Nixon and Emma Carroll bring the joy of youth to the lab and remind me what being young is like. I will not mention Brendan Maguire in my acknowledgements. Finally, I have to thank Susan Marqusee for being a great mentor, boss and friend. I could not have done any of this without her understanding, brilliance or knowledge.

Finally, I have to thank Rosalie Lawrence. She has been such a stabilizing influence in my life and I could not have done this without her patience, understanding, love and support.

Thank you.

Table of Contents

Chapter 1
Introduction: Protein folding *in vivo* and *in vitro*

1.1 Protein folding *in vivo* and *in vitro*

Proteins perform the majority of the chemistry necessary for life on earth and play diverse roles in the cell, from making up the cell's cytoskeleton to being the principal catalysts for energy generation. Proteins are the primary catalysts for DNA replication, signal transduction and cellular metabolism. Most proteins need to adopt a specific three-dimensional structure in order to carry out these diverse functions and modulating protein function by targeting specific aspects of their structure is a specific goal and challenge of modern medicine. In order to understand and predict how the interplay of proteins and their functions constitute, with other biomolecules, the function of a cell, much less an organism, we require rigorous quantitative descriptions of a protein's conformation and behavior. This description is often referred to as a protein's energy landscape. Most energy landscapes are depicted as funnel shaped[1] **(Figure 1.1)**, as most proteins fold from a high energy, heterogeneous unfolded ensemble to a lower energy native state. It is within this native-well that most proteins function.

Anfinsen[2] demonstrated that a protein's entire energy landscape can be encoded by the information in its primary sequence in 1959. Since then, there has been incredible progress made to understand protein folding from a physical point of view. Historically, these types of protein folding experiments are done in conditions suited for protein spectroscopic methods, such as circular dichroism, fluorescence, and NMR. Thus, most experiments that seek to understand folding by measuring proteins' physical chemical properties have been performed *in vitro* under conditions which often do not recapitulate those *in vivo*. In addition, the target proteins for these studies must "behave" properly during those experiments, limiting the number of proteins that can be studied. Despite these disadvantages, we have learned a huge amount about protein folding and the results have played a large role in informing general protein function as well as providing useful information for understanding protein folding simulation and theory. Simulation, too, has similar restrictions (such as protein size and simulation time) and the subset of proteins studied is not totally representative of those in the cell at any given time. Still, folding experiments and simulations have realized many general, quantitative principles that are necessary for understanding the inner workings of the cell.

In the cell, proteins undergo a lifecycle: they are synthesized by the ribosome, fold either during translation or after release, are subsequently released into their functional environment, and finally can be degraded by the cell's proteolytic machinery. There is always the danger, however, of protein misfolding. Buildup of misfolded products can lead to the overwhelming of cellular quality control machinery, which in turn feeds back into the regulation of protein synthesis, maturation, folding and degradation. The presence of protein misfolding and the misregulation of the proteostatic machinery is a hallmark of certain degenerative diseases, such as Alzheimer's and ALS, and amyloids are known to be causative agents for many human diseases[3,4]. Understanding the

physical chemistry of proteins during their cellular life cycle (whether *in vivo* or in reconstituted systems) will allow scientists to better predict protein structures from sequence, what and by what mechanism mutation causes disease, and understand the physical basis for cellular and organismal function.

1.2 *in vivo* biophysical studies have revealed that kinetic folding processes are more likely to be altered than those at equilibrium

In the past 15 years, a few groups have begun the challenge of measuring physical properties of proteins *in vivo.* The cellular environment differs greatly than those in most traditional *in vitro* folding experiments: protein concentrations in the cell are higher than 300g/L and so macromolecular crowding may change how a protein behaves versus *in vitro*[5]. In addition, protein chaperones are constantly surveilling the proteome in order to minimize misfolding and recruit misfolded proteins for degradation. There is some disagreement, however, about how the properties of proteins change *in vivo*. Early studies using hydrogen exchange mass spectrometry (HX-MS) showed a very slight increase in the stability of the lambda repressor N-terminal domain in live cells[6] versus lysate or *in vitro*. These studies subject cells to high concentrations of urea during the pulse and so actions due to chaperones and molecular crowding may be different than in intact cells in a more osmotically favorable environment. In-cell NMR studies of SH3 domain and temperature-jump FRET studies of yeast-phospho-glycerate kinase (PGK)[8,9] both reveal a slight increase in *in vivo* stability similar to those studies using HX-MS. In cell NMR reveals that this increase is recapitulated only by crowded solutions of charged molecules, contrary to earlier *in vitro* studies suggestion that large, neutral crowders are good model systems for cytoplasmic crowding. The aforementioned PGK studies do observe both slower folding and a decrease in apparent cooperativity in-cells compared to experiments undertaken *in vitro*. In contrast, the stability of an engineered protein containing a FlAsH-tag (which binds biarsenical flouresceine) was measured *in vivo.* These experiments, again, were performed by subjecting *E. coli* cells to different concentrations of urea and reading out the fluorescence intensity at each concentration. The investigators observed a significant decrease in the protein's *in vivo* stability[7]. Further experiments undertaken on a different yeast-PGK system than mentioned above, but by the same group, showed a significant increase in PGK stability but also more rapid folding *in vivo*[10]. A caveat of these experiments is that PGK substrate concentrations are at or near their K_m in the cell and not in the *in vitro* experiments performed by those groups,[11]and the protein is not necessarily sampling its unfolded state but more likely a smaller conformational change that leads to loss of FRET. It seems that, at first approximation, proteins have similar physical properties (stability and folding/unfolding rates) *in vivo* and *in vitro* and that how proteins respond to *in vivo* versus *in vitro* conditions will be dependent on the protein itself. There are still many more experiments to be done in order to understand how the environment *in vivo* changes the physical properties of proteins. Furthermore, it is unclear from these studies above how a protein's synthesis, remodeling and modification affect its *in vivo* properties.

1.3 Nascent chains are under heavy proteostatic surveillance and there are specific ribosome-associated chaperones that are necessary for cell-survival and function

Most traditional protein-folding experiments are conducted by unfolding a protein using a denaturant (such as urea, guanidine, temperature or force) and then refolding the protein by dilution. The result of this is that the protein has its entire sequence available for refolding, a situation which may not exist during protein synthesis. Proteins are translated between 2 (eukaryotes) and 10 (prokaryotes) amino acids per second[12,13]. This allows ample time for conformational rearrangements during translation in situations where much of the protein either is occluded (by the ribosome exit tunnel) or not yet synthesized. It is not unreasonable to suggest that these rearrangements may change the folding trajectory of the protein. Protein chaperones fulfill a similar role in the cell by actively remodeling proteins, binding to regions in order to occlude the rest of the sequence during folding, or providing a folding surface that would otherwise be unavailable. It is estimated that about thirty percent of the proteome in *E. coli* requires protein chaperones in order to fold[14]. Translation and chaperone action are extensively remodeled during cellular stress, highlighting the important role that both processes play in contributing to cellular proteostasis[15–17]. A hallmark of the proteotoxic stress response in cells is translational pausing and ribosomal stalling as well as changes in chaperone distribution to actively translating ribosomes[15,16]. One example of this is that translational stalling is mediated by the absence of ribosome-associated chaperones. These chaperones leave the ribosome in response to large numbers of misfolded proteins in the cytosol due to proteotoxic stress (such as temperature or reducing agents). Furthermore, several experiments have shown that inhibiting translation can hinder the formation of toxic aggregates[17]. Indeed, the misregulation of translation and chaperones can contribute to aging[18–20] and the ribosome is a regulatory hub for dealing with proteotoxic stress[21]. It is worth understanding the roles of ribosome-associated chaperones and their mechanisms of action as they may also give hints into how folding occurs during translation.

The eukaryotic quality control machinery is made up of three chaperone containing complexes: the Ribosome-Associated Complex (RAC), the Nascent polypeptide-Associated Complex (NAC) and the Ribosome Quality Control Complex (RQC). The RQC recognizes stalled 80S ribosomes and tags them for nascent chain degradation and ribosome recycling using a novel mechanism involving non-template mediated peptidyl-transfer (so called CAT-tails)[22,23]. Failure of RQC machinery can cause aggregation and proteotoxicity[24].

RAC is the most well-characterized of the three eukaryotic ribosome associated quality control complexes and is a dimer composed of an HSP70 homologue and a J-domain protein. Generally, J-proteins provide specificity to HSP70-type chaperones. Surprisingly, RAC does not make contact with the nascent chain but instead controls the substrate specificity of another Hsp70 protein called SSB. The role of HSP70s in protein folding and remodeling is extensively characterized[25,26] and SSB plays a similar

role as classic HSP70s, but specifically at the site of protein synthesis. RAC also stabilizes specific conformations of the ribosome itself, lending support to the idea that the ribosome also can take an active role in modulating protein folding[27].

NAC, on the other hand, has a more mysterious function but it is known to occupy the same ribosomal binding site as the signal recognizition particle (SRP) and thus deficiencies in NAC can lead to widespread proteotoxic effects due to mis-targeting of proteins not meant for the endoplasmic reticulum. NAC knock-outs are lethal in mice, flies and worms but not in yeast[28]. NAC is thought to function similarly to the prokaryotic chaperone trigger factor (TF), which binds nascent chains so folding may proceed only after enough sequence has been synthesized[29].

While *E. coli* (and other prokaryotes) has many nascent chain modifying enzymes (such as methionyl-aminopeptidase), the repertoire of chaperones that are known to engage with nascent chains is limited primarily to Trigger Factor (TF)[29]. TF is an ATP-independent chaperone composed of three domains: (1) a peptidyl-proline isomerase (PPI) domain; (2) a chaperone domain and (3) a ribosome binding domain **(Figure 1.2)**. It engages nascent chains after about 100 amino acids have been translated (60 amino acids outside of the ribosome exit tunnel) and promiscuously interacts with much of the cytoplasmic proteome[30]. TF engagement, however, is highly enriched for certain proteins classes, especially for outer membrane proteins as well as ß-barrels[30]. TF primarily ensures nascent chain folding fidelity by binding the nascent chain and holding it in a non-native, unfolded-like conformation and nascent chain folding within the ribosome exit tunnel can affect TF recruitment[31]. TF-bound nascent chains are more susceptible to proteolysis and the force exerted by TF-bound nascent chain folding is reduced[32,33]. Single molecule studies of MBP in the presence and absence of TF reveal that TF stabilizes heterogeneous intermediate-like structures[34] and helps prevent aggregation of MBP truncations. This is emblematic of an emergent theme of co-translational protein folding: it is important that intramolecular nascent chain interactions do not result in permanent misfolds before the entire sequence of the nascent protein or domain emerges from the ribosome. Protein chaperones at the ribosome clearly play an important role in the folding of cellular proteins. Nevertheless, it is clear that the vast majority of proteins made by the cell do not require chaperone activity in order to reach their native state. If this is true, however, why is it so difficult to express native proteins recombinantly in different organisms? Why are some proteins able to be expressed recombinantly, but not able to refold even if they do not require chaperone activity in the first place? Obviously, translation and chaperones are different in different organisms and there are many factors that work in concert to answer these questions[35]. It is clear, however, that the ribosome can change a protein's folding efficiency by simply occluding un-synthesized sequence from the exposed nascent chain **(Figure 1.3)**.

1.4 Co-translational folding modulates folding efficiency

Since Anfinsen's original experiments, scientists have wondered how the protein synthesis process could change protein folding. A few studies have been performed with protein fragments to mimic the effect of the nascent chain emerging from the

ribosome, but it is difficult to resolve whether these fragment studies recapitulate the vectorial translation process[36,37]. Several landmark studies, however, have shown that translation does modulate the folding efficiency and rate of protein folding even using *in vitro* extract-based translation systems. The refolding of firefly luciferase is known both to require chaperones and to be slower than folding of luciferase after translation[38,39]. The faster folding rate and higher folding efficiency of firefly luciferase during translation is due to the formation of a ribosome-obligate folding intermediate[40]. Similarly, P22 tailspike (a protein that mediates the recognition of P22 phage with *S. enterica* cells) folding also proceeds through several translation dependent intermediates that are not present during refolding and results in a much more efficient protein folding pathway[41]. Some of these intermediates can be trapped and are present at equilibrium on truncated, stalled, nascent chains but upon release most of these chains form aggregates. The protein Flavodoxin also adopts an intermediate (in this case an off-pathway, molten globule-like intermediate) during refolding that, in contrast, drastically slows its observed folding rate. However, population of this intermediate is abrogated during co-translational folding[42]. This situation is similar to bacterial luciferase, which is an alpha/beta heterodimer that assembles more rapidly and efficiently *de novo* (i.e. during translation) versus during refolding. Co-translational folding allows the beta subunit of bacterial luciferase to avoid a long-lived intermediate that is populated only during refolding[43]. Skipping this slow kinetic step, however, is at least in-part due to the presence of excess alpha subunit. When the beta subunit is translated in the absence of the alpha subunit, its folding proceeds through the slow folding step observed during refolding. This phenomenon – that quarternary structure formation can be facilitated by co-translational folding – has also been observed *in vivo*. Selective ribosome profiling studies have shown that complex assembly is influenced by operon construction and perhaps subunit assembly is tuned to when other subunits are emerging from the ribosome[44]. Finally, GFP folding after ribosomal release is also highly efficient (75% from stalled ribosomes compared to 10% during refolding)[45] and this efficiency difference is due to formation of a co-translational obligate folding intermediate. It is interesting to note that every protein mentioned here contains a large fraction of ß-sheets, has a somewhat complex topology, (especially in the case of P22 tailspike; **Figure 1.4**) and is larger than the average protein used for *in vitro* protein folding studies[46].

It is clear that the vectorial nature of translation allows the nascent chain time to collapse before problematic segments of the protein emerge from the ribosome. This plays an important role in shepherding the nascent chain on the most efficient route to the native state. Translation is not, however, a continuous process and thus the rate of translation may help provide time for certain segments of the protein to refold before other segments have been synthesized.

1.5 Translation rate modulates co-translational folding

Translation rate depends on a multitude of conditions including codon bias, available peptidyl-tRNA, cofactor occupancy and availability (such as elongation factor-

P[47,48]), and direct regulation. Since the discovery that synonymous codons (codons that code for the same amino acid but are different in sequence) and their cognate tRNA genes occur at different frequencies throughout the genome, scientists have had a range of hypotheses for why this would occur and what the selective pressures for it would be[12,49–51]. Furthermore, in the last two decades, researchers have determined that synonymous mutations are associated with a wide range of diseases[52] and there is evidence that codon bias is related to mis-translation, gene expression levels and specific cellular processes that dominate sequence selection[53–55]. It is clear that codon choice does influence translation rate and that translation rate and accuracy are under selective pressure. The hypothesized connection between translation rate (and thus codon usage) and protein folding has been under investigation for some time. Perhaps proteins translate more slowly after single domains have emerged in multi-domain proteins, or codons for slow forming secondary structures are followed by slow translating codons. Contrary to this hypothesis, it has been difficult to find correlations between secondary structure formation and codon bias even with large data sets[51,55–57]. Computational and theoretical studies, however, have shown that changes in codon usage can change translation rate and that there are weak signatures of a relationship between codon usage and folding of certain secondary structural elements[58–61]. It is important to note that normalizing codon usage by measured tRNA concentrations reveals stronger correlations than just tRNA gene copy number[58,62] when performing these calculations. It is only recently, however, that experimental proof of the relationship between translation rate and the physical properties of proteins has been understood.

Since it is difficult to assess the rate of folding in an asynchronous process, most investigations of the translation-speed/protein folding question are performed on proteins with a functional read-out. This means that these types of experiments measure drastic changes in a protein's structure in relation to a changing translation rate and these changes must be trapped after ribosomal release. Clark and colleagues used a highly engineered split-GFP construct to show that translating quickly through an artificial linker could bias the final population of GFP[63]. They used three split GFP domains, a constant region (K) and either a blue light emitting fragment (B) or yellow light emitting fragment (Y). When the linker between B and Y was translated quickly, a higher proportion of KB was formed, leading to an increase in the B/Y emission ratio. This emission ratio was titratable depending on the translation speed of this linker. Similarly, three naturally occurring synonymous mutations result in altered function of an ATP-driven efflux pump called Pg-p encoded by the human Multi-Drug Resistance 1 gene (MDR1)[64]. Even though protein levels remain constant in the presence of these synonymous mutations, the maturation of Pg-p is affected, as Pg-p is more susceptible to limited proteolysis. Additionally, Pg-p's substrate specificity is altered depending on which silent mutation is present. The presence of substrate, however, suppresses mutant Pg-p trypsin susceptibility, suggesting that the synonymous codon usage alters the conformational equilibrium of Pg-p, but the protein can be remodeled to the WT conformation. Similar experiments investigating the membrane protein Fus1 uncovered a similar effect. Changes to the translation speed of Fus1 altered both its susceptibility

to proteolysis and the buildup of protease-protected co-translational intermediates[65]. Finally, a study on Gamma-B-crystallin showed that its native translation rate is necessary for proper disulfide formation and that if disulfide scrambling occurs (due to altered translation rates), the protein adopts a different conformation after release than if translation rate were not altered[66]. These studies highlight the connection between co-translational folding and translation rate as well as the current limitations of the field.

For many proteins, translation speed/codon usage only makes a difference in terms of expression or solubility[67], but the mechanism by which these many of these differences manifest themselves remain a mystery. This is due to the fact that obtaining a mechanistic view of protein folding during translation *in vitro* or *in vivo* is experimentally challenging. So, most groups measure conformational differences indirectly (for instance by assessing protein function) or after the protein has already reached equilibrium. This limitation has biased the field either towards highly engineered folding substrates or towards membrane proteins whose insertion in the membrane has long been known to be modulated by mRNA sequence and translation speed[57]. Whether the changes seen in the processing, function and conformations of membrane proteins are *bona fide* protein folding issues or issues of insertion, trafficking or chaperone engagement is currently under debate. Complicating these analyses, especially when comparing *de novo* folding to refolding, is the fact that the translational machinery, and especially the ribosome, can affect the nascent chain via direct interactions.

1.6 Interactions of the nascent chain with the ribosome

For over fifty years, scientists have known that the ribosome protects regions of the nascent chain from proteolysis[68] and that many antibiotics target the ribosome exit tunnel by blocking the nascent chain from emerging. It is only within the past fifteen years, however, that the surface of the ribosome and especially the ribosome exit tunnel has become a target for researchers trying to understand co-translational folding. In *E. coli*, the ribosome exit tunnel is about 100Å long and two ribosomal proteins, L22 and L4 protrude into it **(See Figure 1.5)**[69]. At the very far end of the channel, near the surface of the ribosome, is what is called the tunnel vestibule, a widening of the tunnel that occurs just after its most constricted point. The tunnel, if the nascent chain were fully elongated, could fit between 30 and 40 residues within it. Furthermore, the ribosome exit tunnel is wide enough, at 13.7Å at its narrowest point, to accommodate the folding of an alpha helix (helices are ~9.5Å in diameter). One of the first observations for a functional role of the ribosome exit tunnel and the peptide within it is the SecM stalling sequence, which stalls the ribosome through interactions between the exit tunnel and the nascent chain[70] and which is now used for generating homogenous ribosome nascent chain complexes (RNCs)[71]. It is now recognized that the tunnel can play diverse functional roles, even acting as a tryptophan sensor[72,73]. Furthermore, it has recently become clear that the nascent chain is able to acquire secondary or even tertiary structure even within the exit tunnel. The presence of helices and ß-turns within the tunnel was first observed using PEG-ylation of the emerging chain in some of the

first studies to make a direct observation of protein collapse in the exit tunnel[74–77]. Later cryo-EM studies then provided the first high-resolution images of alpha-helical structure formation within the ribosome exit tunnel[78].

One major unanswered question is how the forces that drive protein folding are balanced when the protein resides in the ribosome exit tunnel. The exit tunnel is permeable to water and certain ions[69,79], but is made-up of highly negatively charged rRNA and, as mentioned above, ribosomal proteins L22 and L4. Recently, cryo-EM structures of a spectrin fragment and a zinc-binding domain from the transcription factor ADR1 have been solved with the folded protein residing within the exit tunnel's folding vestibule[80,81]. Unfortunately, there are no clear trends that relate the physical properties of proteins (studied *in vitro*) to structure acquisition in the tunnel (and during translation)[80]. This may be because the nascent chain can interact with the surface of the ribosome, which NMR and dynamic fluorescence depolarization (DFP) studies have suggested[82,83]. These DFP studies stalled intrinsically disordered protein fragments as RNCs and systematically changed the charge on the exposed peptide. They found that positively charged peptides displayed a reduced range of motion compared to peptides with a negative charge and compared to some peptides off the ribosome. Interestingly, they also found that ribosome-stalled peptide fragments behave similarly to those same fragments in denaturing conditions, suggesting that solution conditions proximal to the surface of the ribosome are different from those after release.

The ribosome exit tunnel and surface are not inert to the emerging chain, as once thought. How do these properties, intrinsic to the ribosome, change nascent chain folding as compared to refolding? Are there general folding phenomena that occur on the ribosome that do not occur during refolding or at equilibirium?

1.7 Co-translational folding promotes the formation of folding intermediates

As discussed above, there is evidence that co-translational folding can significantly change protein function and folding efficiency. A common hypothesis to explain this is that the slow, vectorial nature of translation allows for rearrangement of the nascent chain during synthesis so that it populates intermediates not allowed during refolding of the protein[84]. So, what evidence is there for this hypothesis? How much does the nascent chain collapse during translation? Two studies published at the turn of the millennium provided a first glimpse into this phenomenon. One described the co-translational formation of a fully functional viral protease domain of a larger viral poly-protein in both eukaryotic and bacterial cells[85]. These results are important because most co-translational folding experiments are done using *in vitro* translation systems, which often translate proteins an order of magnitude more slowly than *in vivo*. Thus, this paper establishes that co-translational rearrangements can occur *in vivo*. The other paper, by Frydman et al. showed that firefly luciferase forms an obligate co-translational intermediate that is necessary for fast folding (discussed in 1.4)[40]. Since then, only a few obligate co-translational intermediates have been observed. P22 tailspike protein has been shown to populate several different stable structures at equilibrium during translation[41]. The mechanism for changes in GFP folding efficiency, originally observed by the Clark lab, were uncovered in a 2012 paper by Skatch and colleagues[86]. This

clever study took advantage of the fact that urea inhibits *de novo* GFP folding, but does not inhibit chromophore formation. Thus, the authors were able to determine the amount of folded protein as a function of nascent chain length as well as the amount of folded protein after release from the ribosome. The observed folding of released GFP, as well as the fact that there was a chymotrypsin-protected fragment present on RNCs provide evidence for intermediate formation during translation. However, these studies do not pinpoint which regions of the protein are involved in intermediate formation. Building on a classic paper where the co-translational folding of an N-terminal ß-galactosidase fragment was tracked using antibodies[87], the co-translational folding of Influenza A hemagglutinin was tracked using "FactSeq."[88] FactSeq is similar to ChIP: ribosomes are converted into monosomes with RNase treatment and antibodies pulldown specific folded epitopes on hemagglutinin. Then, the mRNA is deep sequenced and the point at which the folded epitope appears can be localized on the mRNA. Thus, the formation of the folded epitope recognized by each antibody can be pinpointed by an increase in the number of reads mapping to a specific region on that mRNA. In order to fully understand the role of the ribosome in nascent chain folding, is essential to have a high resolution structural view of RNCs.

1.8 Investigations into the structure of ribosome nascent chains

Researchers have tried for years to obtain high resolution structures of nascent chains. Unfortunately, nascent chains are highly dynamic and attached to a macromolecular machine a few orders of magnitude more massive. A major breakthrough in the study of RNCs is the specific labelling and monitoring of nascent chains by NMR[83,89–92]. These studies have revealed that there is a "folding offset" for the protein ddFLN5 (an IgG like domain from the amoeba Dictyostelium discoideum) during translation– even when the sequence required for folding has emerged from the ribosome exit tunnel, folded-like resonances have not yet reached their full intensity. In addition, these studies revealed large numbers of transient interactions between the linker used in the studies and the ribosomal surface, buttressing arguments that the ribosomal surface is an important factor in determining co-translational folding. Similar nascent chain dynamics were observed for an SH3 domain[89]. Real-time observation of the nascent chain during translation using FRET has also been carried out. These experiments involve the purification of stalled RNCs, whose stalling is then relieved by the addition of the donor fluorophore using a stopped-flow mixing apparatus. These studies reveal collapse of the nascent chain before the full sequence is out of the tunnel[93]. Complementary to these studies are single-molecule pulling studies on T4-lysozyme, which show a decrease in folding rate that is dependent on distance from the peptidyl-transferase center (the closer to the PTC, the greater the magnitude of the folding rate decrease)[94]. Interestingly, this decrease in folding rate is partially attenuated by an increase in the ion strength of the solution, supporting the hypothesis that folding is modulated by the charge of the ribosomal surface. Finally, some time should be spent understanding a new type of assay that has been introduced by the Von Heijne group[81]. This assay measures read-through of a moderately-weak stalling

sequence due to the force generated by folding of the nascent chain. By varying the linker length between the stall sequence and the protein, a force-length profile is measured. This assay has a characteristic bell-shape. When the nascent chain is too close to the PTC, it is sterically restricted and thus cannot fold; when it is far away, the force is not relayed to the PTC and thus read through does not occur. This assay has been used to profile a large number of proteins and has shown that the force generated on the stall sequence can be modulated by trigger factor as well as cofactor binding[32]. However, there is no definitive study that shows the mechanism by which force is relayed to the stall sequence nor is the force correlated with the folding pathway, stability or folding kinetics of the nascent chain[80].

1.9 Where is co-translational folding as a field and where is it going?

The folding field in general has started moving in two different directions at once. First, it is moving towards small timescales and atomic resolution by increasing the resolution if its experiments– ultra-fast mixing experiments using microfluidics and single amino acid resolution kinetic folding experiments using HXMS and other methods (such as smFRET or NMR). Second, the folding field is moving towards understanding the role of protein folding in the context of more complicated systems– moving towards measuring folding during translation and *in vivo*, using more powerful algorithms and computers to perform folding simulations, and layering probes on top of each other (for instance, combining two types of denaturants and two types of measuring protein folding). As this continues, recent advances in the creation and manipulation of RNCs have allowed for a more quantitative view of protein folding to develop. However, the co-translational folding field lags far behind in terms of the measurement of biophysical properties that can be used for benchmarking simulation experiments and theory. Part of the reason for this is that those studying co-translational folding have historically been biochemists and molecular biologists and, in general, those studying folding traditionally come from a chemistry or physics background.

There are a few major hurdles that need to be surpassed in order for co-translational folding to be understood at the level of general protein folding: (1) New experimental or mathematical modelling methods for synchronizing co-translational folding and translation. That is, a major challenge is that translation reactions are highly asynchronous and so the measurement of folding rate is convoluted with the synthesis of the protein itself, thus making the measurement of folding rates impossible. (2) An increase in throughput for measuring biophysical properties of nascent chains. Currently, the most quantitative, biophysical studies available use techniques such as single-molecule force spectroscopy, FRET, or NMR, all of which are limited by variables such as attachment sites and protein behavior and yield. (3) General probes for distinguishing the nascent chain from the ribosome. Too often, protein function is used as a read out for folding and so certain techniques are not applicable in a general way, limiting the usefulness of their conclusions. (4) Understanding how the ribosome responds to common reagents for accessing high energy states (such as chemical denaturants). (5) Verifiable theoretical models for co-translational folding.

The time is ripe for collaboration between these two communities and we have seen some papers published jointly from members of both communities (for instance, [80]). Understanding co-translational folding may play a key role in solving the folding problem, as it is possible that translation places severe restrictions on the folding of most proteins and thus could help restrict *in silico* folding space.

1.10 Summary of work contained in this dissertation

Here, I introduce a new method, pulse proteolysis, for probing both the stability ($\Delta G_{unfolding}$) and unfolding rates of ribosome nascent chains. Included in this work, I assess how the ribosome responds to the denaturant urea and show that stalled ribosomes are fundamentally more urea-resistant in comparison to un-stalled 70S ribosomes. We find that the ribosome destabilizes nascent chains and that this destabilization primarily is the result of an increase in the rate of unfolding, rather than a slowing of the folding rate. This destabilization is dependent on the isoelectric point (pI) of the emerging chain, which is consistent with the idea that the high charge density of the ribosomal surface modulates nascent chain folding. Next, we adapt this pulse proteolysis method to measure the unfolding rates of two proteins: DHFR V75R (a destabilized variant of DHFR) and RNase H I53D (a two-state variant of RNase H). For DHFR V75R, we observe slight acceleration in the unfolding rate from one of its unfolding intermediates. In contrast, for RNase H I53D, we conclude that the ribosome changes RNase H's unfolding rate more than its folding rate, but does not alter its kinetic m-value. Finally, I use the protein Halo-tag to develop a high-throughput method for measuring co-translational folding and determine the folding pathway of Halo-tag using hydrogen exchange mass-spectrometry (HXMS). Using this information, I investigate the folding pathway of Halo-tag during translation and conclude that the ribosome abrogates the formation of an early-forming Halo-tag folding intermediate that causes aggregation. The alteration of Halo-tag's folding pathway, which only happens during *de novo* folding, significantly increases the folding efficiency of Halo-tag. First however, I would like to explain the principles and limitations of pulse proteolysis.

1.11 Pulse proteolysis is a gel-based method for measuring the fraction of folded protein in a sample

Pulse proteolysis was originally conceived as a gel-based method for measuring protein stability[95,96]. Pulse proteolysis relies on the fact that unfolded proteins are more susceptible to proteolysis than folded proteins. This is because the conformational flexibility of unfolded proteins is greater than that of folded proteins. Additionally, unfolded proteins have a greater percentage of their sequence available for proteolysis. So, by subjecting a protein at equilibrium to a short pulse of protease and immediately quenching, all the unfolded protein in a given sample can be cleaved while the folded protein remains intact. This process is repeated at varying concentrations of denaturant and then each reaction is run on a gel. The intensity of the full-length band is directly proportional to the amount of folded protein at that given denaturant concentration.

Quantitation of the full-length band intensities can then be plotted against the urea concentration and fit to the equation below to yield the C_m, or the denaturant midpoint:

$$\text{Eq(1): } I = I_0(1/(1+\exp(-\Delta G_{unf}/RT)); \ \Delta G_{unf} = -m(C_m - [\text{urea}])$$

When fitting, I_0 should be fixed to the intensity of no thermolysin sample (however, in cases where there is a sub-population of protein that does not exchange with the folded and unfolded states, it is okay to determine I_0 by fitting it to a model). The C_m is then multiplied by the experimentally determined or theoretically calculated m-value to determine the stability. Pulse proteolysis currently is too noisy to determine m-values accurately. In addition, it is unclear whether proteolysis m-values are the same as m-values obtained using traditional methods. It is possible, however, that by combining many high-quality datasets, or by obtaining a large number of points in the unfolding transition region, that accurate m-values can be measured using pulse proteolysis.

The denaturant used during pulse-proteolysis must be compatible with the method of reading out the amount of full-length protein. Therefore, urea is most often used as a denaturant. However, Guanidine Hydrochloride can be used as long as the protein is TCA precipitated before loading on a gel. High concentrations of salt cause gels to run poorly. Other denaturants, such as temperature or SDS have been used successfully with pulse proteolysis[97–99].

It is important that (1) the pulse is long enough that all the unfolded protein is cleaved, (2) that the unfolding rate of the target protein is long enough that a negligible amount of protein unfolds during the pulse time and (3) the native state of the target protein does populate excited states that are protease susceptible. In practice, the first caveat is more of a problem than the second or third. The protease must be active enough in high denaturant conditions in order to cleave all unfolded protein (especially for very stable proteins). In addition, quenching must be irreversible and occur very quickly compared to the pulse length. For this reason, the thermostable protease thermolysin has been traditionally used. Thermolysin retains its activity in high urea and can be quenched by the addition of EDTA, which is an irreversible quenching step. Thermolysin is also very sensitive to reducing agents and requires 10mM $CaCl_2$ to function, especially in highly denaturing conditions. For these reasons, other proteases are sometimes used, including Subtilisin[98]. Alternatively, the concentration of thermolysin can be increased or decreased to tune the proteolysis rate. It is extremely important, however to be sure that the experiment remains in the kinetic regime appropriate for pulse proteolysis to accurately report on the stability of the target protein. Future work may want to look into using kinetically trapped proteases, such as alpha-lytic protease, for use in pulse proteolysis.

In order to determine if a protein's proteolysis and folding kinetics are suitable for pulse proteolysis, two experiments should be undertaken. The first is that the proteolysis rate of the folded and unfolded states should be measured near the C_m of the protein. The second (especially if the Cm is unknown), is to run a zero-urea sample with and without thermolysin treatment. If there is any cleavage of the native state at zero molar urea, then the protein either experiences deviations away from its native state during the pulse time or its unfolding rate is so rapid that it is unsuitable for pulse

proteolysis.

Pulse proteolysis may be used to determine unfolding and folding rates[100] and dissociation constants (K_d) of proteins and their ligands. One caveat when measuring folding rates is that it is especially important to run a no thermolysin sample in order to measure the total amount of protein that is refolded. Kinetic data is fit to an exponential and then plotted as a function of urea. A practical note when performing kinetic experiments is that each sample should be assayed one minute before the time point in question, as unfolding or refolding continues during the pulse.

Pulse proteolysis is a diverse method that can be applied not just for determining protein stability and has been used to simply assess the fraction of folded protein in different conditions, including during translation[93,101]. If the protein is labelled and accessible, it can be applied in heterogeneous solutions, making it somewhat more flexible than traditional protein spectroscopy. In addition, pulse proteolysis has the advantage of using miniscule amounts of protein, is applicable in diverse buffer conditions, and uses standard benchtop lab equipment. In sum, pulse proteolysis is a great method for measuring biophysical properties of proteins that are not compatible with traditional spectroscopic methods.

1.12 References

1. Dill K a, MacCallum JL. The protein-folding problem, 50 years on. *Science*. 2012;338(6110):1042-1046. doi:10.1126/science.1219021.
2. HABER E, ANFINSEN CB. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem*. 1961;236(2):422-424. http://academygenbioii.pbworks.com/f/Chapter+3+investigation.pdf%5Cnfiles/14635/1961_Haber_J Biol Chem.pdf.
3. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426(6968):884-890. doi:10.1038/nature02261.
4. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591.
5. Albe KR, Butler MH, Wright BE. Cellular concentrations of enzymes and their substrates. *J Theor Biol*. 1990;143(2):163-195. doi:10.1016/S0022-5193(05)80266-8.
6. Ghaemmaghami S, Oas TG. Quantitative protein stability measurement in vivo. *Nat Struct Biol*. 2001;8(10):879-882. doi:10.1038/nsb1001-879.
7. Ignatova Z, Gierasch LM. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci U S A*. 2004;101(2):523-528. doi:10.1073/pnas.0304533101.
8. Smith AE, Zhou LZ, Gorensek AH, Senske M, Pielak GJ. In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci U S A*. 2016;113(7):1725-1730. doi:10.1073/pnas.1518620113.
9. Ebbinghaus S, Dhar A, McDonald JD, Gruebele M. Protein folding stability and dynamics imaged in a living cell. *Nat Methods*. 2010;7(4):319-323. doi:10.1038/nmeth.1435.

10. Guo M, Xu Y, Gruebele M. Temperature dependence of protein folding kinetics in living cells. *Proc Natl Acad Sci U S A*. 2012;109(44):17863-17867. doi:10.1073/pnas.1201797109.
11. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol*. 2009;5(8):593-599. doi:10.1038/nchembio.186.
12. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985;2(1):13-34. http://www.ncbi.nlm.nih.gov/pubmed/3916708.
13. Young R, Bremer H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem J*. 1976;160(2):185-194. doi:10.1042/bj1600185.
14. Kerner MJ, Naylor DJ, Ishihama Y, et al. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell*. 2005;122(2):209-220. doi:10.1016/j.cell.2005.05.028.
15. Shalgi R, Hurt JA, Krykbaeva I, Taipale M, Lindquist S, Burge CB. Widespread Regulation of Translation by Elongation Pausing in Heat Shock. *Mol Cell*. 2013;49(3):439-452. doi:10.1016/j.molcel.2012.11.028.
16. Liu B, Han Y, Qian S-B. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol Cell*. 2013;49(3):453-463. doi:10.1016/j.molcel.2012.12.001.
17. Zhou C, Slaughter BD, Unruh JR, et al. Organelle-based aggregation and retention of damaged proteins in asymmetrically dividing cells. *Cell*. 2014;159(3):530-542. doi:10.1016/j.cell.2014.09.026.
18. Labbadia J, Morimoto RI. The Biology of Proteostasis in Aging and Disease. *Annu Rev Biochem*. 2015;(March):1-30. doi:10.1146/annurev-biochem-060614-033955.
19. Roth DM, Balch WE. Modeling general proteostasis: proteome balance in health and disease. *Curr Opin Cell Biol*. 2011;23(2):126-134. doi:10.1016/j.ceb.2010.11.001.
20. Walther DM, Kasturi P, Zheng M, et al. Widespread Proteome Remodeling and Aggregation in Aging C. elegans. *Cell*. 2015;161(4):919-932. doi:10.1016/j.cell.2015.03.032.
21. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol*. 2009;16(6):589-597. doi:10.1038/nsmb.1614.
22. Shen PS, Park J, Qin Y, et al. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science (80-)*. 2015;347(6217):75-78. doi:10.1126/science.1259724.
23. Brandman O, Stewart-Ornstein J, Wong D, et al. A Ribosome-Bound Quality Control Complex Triggers Degradation of Nascent Peptides and Signals Translation Stress. *Cell*. 2012;151(5):1042-1054. doi:10.1016/j.cell.2012.10.044.
24. Choe Y-J, Park S-H, Hassemer T, et al. Failure of RQC machinery causes protein aggregation and proteotoxic stress. *Nature*. 2016;531(7593):191-195. doi:10.1038/nature16973.
25. Mayer MP. Hsp70 chaperone dynamics and molecular mechanism. *Trends*

Biochem Sci. 2013;38(10):507-514. doi:10.1016/j.tibs.2013.08.001.
26. Mayer MP, Bukau B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci*. 2005;62(6):670-684. doi:10.1007/s00018-004-4464-6.
27. Zhang Y, Ma C, Yuan Y, et al. Structural basis for interaction of a cotranslational chaperone with the eukaryotic ribosome. *Nat Struct Mol Biol*. 2014;21(12). doi:10.1038/nsmb.2908.
28. del Alamo M, Hogan DJ, Pechmann S, Albanese V, Brown PO, Frydman J. Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol*. 2011;9(7). doi:10.1371/journal.pbio.1001100.
29. Pechmann S, Willmund F, Frydman J. The ribosome as a hub for protein quality control. *Mol Cell*. 2013;49(3):411-421. doi:10.1016/j.molcel.2013.01.020.
30. Oh E, Becker AH, Sandikci A, et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*. 2011;147(6):1295-1308. doi:10.1016/j.cell.2011.10.044.
31. Lin K-F, Sun C-S, Huang Y-C, et al. Cotranslational protein folding within the ribosome tunnel influences trigger-factor recruitment. *Biophys J*. 2012;102(12):2818-2827. doi:10.1016/j.bpj.2012.04.048.
32. Nilsson OB, Müller-Lucks A, Kramer G, Bukau B, von Heijne G. Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein. *J Mol Biol*. 2016;428(6):1356-1364. doi:10.1016/j.jmb.2016.02.014.
33. Hoffmann A, Becker AH, Zachmann-Brand B, Deuerling E, Bukau B, Kramer G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol Cell*. 2012;48(1):63-74. doi:10.1016/j.molcel.2012.07.018.
34. Mashaghi A, Kramer G, Bechtluft P, et al. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature*. 2013;500(7460):98-101. doi:10.1038/nature12293.
35. Netzer WJ, Hartl FU. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*. 1997;388(6640):343-349. doi:10.1038/41024.
36. Kippen a D, Sancho J, Fersht a R. Folding of barnase in parts. *Biochemistry*. 1994;33(12):3778-3786.
37. Reymond MT, Dyson HJ, Wright PE, Merutka G. Folding propensities of peptide fragments of myoglobin. *Protein Sci*. 1997;6(3):706-716. doi:10.1002/pro.5560060320.
38. Spirin AS. Folding of firefly luciferase during translation in a cell-free system. 1994;13(15):3631-3637.
39. Kolb VA, Kommer A, Spirin AS. Co-Translational Protein Folding in Prokaryotic and Eukaryotic Cell-Free. 2002;7(95).
40. Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*. 1999;6(7):697-705. doi:10.1038/10754.

41. Evans MS, Sander IM, Clark PL. Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol*. 2008;383(3):683-692. doi:10.1016/j.jmb.2008.07.035.
42. Houwman JA, André E, Westphal AH, van Berkel WJH, van Mierlo CPM. The Ribosome Restrains Molten Globule Formation in Stalled Nascent Flavodoxin. *J Biol Chem*. 2016:jbc.M116.756205. doi:10.1074/jbc.M116.756205.
43. Fedorov AN, Baldwin TO. Process of biosynthetic protein folding determines the rapid formation of native structure. *J Mol Biol*. 1999;294(2):579-586. doi:10.1006/jmbi.1999.3281.
44. Shieh Y-W, Minguez P, Bork P, et al. Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science*. 2015;350(6261):678-680. doi:10.1126/science.aac8171.
45. Ugrinov KG, Clark PL. Cotranslational folding increases GFP folding yield. *Biophys J*. 2010;98(7):1312-1320. doi:10.1016/j.bpj.2009.12.4291.
46. Braselmann E, Chaney JL, Clark PL. Folding the proteome. *Trends Biochem Sci*. 2013;38(7):337-344. doi:10.1016/j.tibs.2013.05.001.
47. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina M V. EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. *Science*. 2012;85. doi:10.1126/science.1229017.
48. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. *Science*. 2012;82. doi:10.1126/science.1228985.
49. Ikemura T, Ozeki H. Codon Usage and Transfer RNA Contents: Organism-specific Codon-choice Patterns in Reference to the Isoacceptor Contents. *Cold Spring Harb Symp Quant Biol*. 1983;47:1087-1097. doi:10.1101/SQB.1983.047.01.123.
50. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol*. 1981;151(3):389-409. doi:10.1016/0022-2836(81)90003-6.
51. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011;12(1):32-42. doi:10.1038/nrg2899.
52. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. 2011;12(10):683-691. doi:10.1038/nrg3051.
53. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008;134(2):341-352. doi:10.1016/j.cell.2008.05.042.
54. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. *Science*. 2009;324:255-258. https://www.pubchase.com/article/19359587.
55. Li G-W, Oh E, Weissman JS. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. March 2012.

doi:10.1038/nature10965.
56. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009;324(5924):218-223. doi:10.1126/science.1168978.
57. Rodnina M V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci*. 2016;25:1390-1406. doi:10.1002/pro.2950.
58. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. December 2012. doi:10.1038/nsmb.2466.
59. O'Brien EP, Vendruscolo M, Dobson CM. Prediction of variable translation rate effects on cotranslational protein folding. *Nat Commun*. 2012;3(May):868. doi:10.1038/ncomms1850.
60. Ciryam P, Morimoto RI, Vendruscolo M, Dobson CM, O'Brien EP. PNAS Plus: In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proc Natl Acad Sci*. December 2012. doi:10.1073/pnas.1213624110.
61. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc*. 2011;133(3):513-526. doi:10.1021/ja107863z.
62. Pechmann S, Chartron JW, Frydman J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol*. 2014;21(12):1100-1105. doi:10.1038/nsmb.2919.
63. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858-861. doi:10.1021/ja411302m.
64. Kimchi-Sarfaty C, Oh JM, Kim I-W, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*. 2007;315(5811):525-528. doi:10.1126/science.1135308.
65. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009;16(3):274-280. doi:10.1038/nsmb.1554.
66. Buhr F, Jha S, Thommen M, et al. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol Cell*. 2016;61(3):341-351. doi:10.1016/j.molcel.2016.01.008.
67. Hess A-K, Saffert P, Liebeton K, Ignatova Z. Optimization of translation profiles enhances protein expression and solubility. *PLoS One*. 2015;10(5):e0127039. doi:10.1371/journal.pone.0127039.
68. Malkin LI, Rich A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol*. 1967;26(2):329-346. http://www.ncbi.nlm.nih.gov/pubmed/4962271.
69. Voss NR, Gerstein M, Steitz T a, Moore PB. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol*. 2006;360(4):893-906. doi:10.1016/j.jmb.2006.05.023.
70. Nakatogawa H, Ito K. The ribosomal exit tunnel functions as a discriminating gate.

Cell. 2002;108(5):629-636. http://biochemistry.ucsf.edu/courses/bioreg/2 star papers 2003/2star3_3_03.pdf. Accessed March 26, 2013.
71. Evans MS, Ugrinov KG, Frese M-A, Clark PL. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods*. 2005;2(10):757-762. doi:10.1038/nmeth790.
72. Bischoff L, Berninghausen O, Beckmann R. Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep*. 2014;9(2):469-475. doi:10.1016/j.celrep.2014.09.011.
73. Lintner NG, McClure KF, Petersen D, et al. Selective stalling of human translation through small-molecule engagement of the ribosome nascent chain. *PLoS Biol*. 2017;15(3):e2001882. doi:10.1371/journal.pbio.2001882.
74. Lu J, Deutsch C. Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol*. 2005;12(12):1123-1129. doi:10.1038/nsmb1021.
75. Lu J, Hua Z, Kobertz WR, Deutsch C. Nascent Peptide side chains induce rearrangements in distinct locations of the ribosomal tunnel. *J Mol Biol*. 2011;411(2):499-510. doi:10.1016/j.jmb.2011.05.038.
76. Tu LW, Deutsch C. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J Mol Biol*. 2010;396(5):1346-1360. doi:10.1016/j.jmb.2009.12.059.
77. Kosolapov A, Deutsch C. Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol*. 2009;16(4):405-411. doi:10.1038/nsmb.1571.
78. Bhushan S, Gartmann M, Halic M, et al. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol*. 2010;17(3):313-317. doi:10.1038/nsmb.1756.
79. Noeske J, Wasserman MR, Terry DS, Altman RB, Blanchard SC, Cate JHD. High-resolution structure of the Escherichia coli ribosome. *Nat Struct Mol Biol*. 2015;22(4):336-341. doi:10.1038/nsmb.2994.
80. Nilsson OB, Nickson AA, Hollins JJ, et al. Cotranslational folding of spectrin domains via partially structured states. *Nat Struct & Mol Biol*. 2017;24(3):221-226. doi:10.1038/nsmb.3355.
81. Nilsson OB, Hedman R, Marino J, et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep*. 2015;12(10):1533-1540. doi:10.1016/j.celrep.2015.07.065.
82. Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, Cavagnero S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem Biol*. 2013;8(6):1195-1204. doi:10.1021/cb400030n.
83. Cabrita LD, Cassaignau AME, Launay HMM, et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol*. 2016;(February). doi:10.1038/nsmb.3182.
84. Evans MS, Clarke TF, Clark PL. Conformations of co-translational folding intermediates. *Protein Pept Lett*. 2005;12(2):189-195. http://www.ncbi.nlm.nih.gov/pubmed/15723645.
85. Nicola a V, Chen W, Helenius a. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol*. 1999;1(6):341-345. doi:10.1038/14032.

86. Kelkar D a, Khushoo A, Yang Z, Skach WR. Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. *J Biol Chem*. 2012;287(4):2568-2578. doi:10.1074/jbc.M111.318766.
87. Hamlin J, Zabin I. -Galactosidase: immunological activity of ribosome-bound, growing polypeptide chains. *Proc Natl Acad Sci U S A*. 1972;69(2):412-416. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=426469&tool=pmcentrez&rendertype=abstract.
88. Han Y, David a., Liu B, et al. Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc Natl Acad Sci*. July 2012:6-11. doi:10.1073/pnas.1208138109.
89. Eichmann C, Preissler S, Riek R, Deuerling E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2010;107(20):9111-9116. doi:10.1073/pnas.0914300107.
90. Hsu S-TD, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2007;104(42):16516-16521. doi:10.1073/pnas.0704664104.
91. Hsu S-TD, Cabrita LD, Fucini P, Christodoulou J, Dobson CM. Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J Am Chem Soc*. 2009;131(24):8366-8367. doi:10.1021/ja902778n.
92. Cabrita LD, Hsu S-TD, Launay H, Dobson CM, Christodoulou J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2009;106(52):22239-22244. doi:10.1073/pnas.0903750106.
93. Holtkamp W, Kokic G, Jäger M, Mittelstaet J, Komar AA, Rodnina M V. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104-1107. doi:10.1126/science.aad0344.
94. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C. The ribosome modulates nascent protein folding. *Science*. 2011;334(6063):1723-1727. doi:10.1126/science.1209740.
95. Park C, Marqusee S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*. 2005;2(3):207-212. doi:10.1038/nmeth740.
96. Park C, Marqusee S. Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr Protoc Protein Sci*. 2006;Chapter 20:Unit 20.11. doi:10.1002/0471140864.ps2011s46.
97. Okada J, Koga Y, Takano K, Kanaya S. Slow unfolding pathway of hyperthermophilic Tk-RNase H2 examined by pulse proteolysis using the stable protease Tk-subtilisin. *Biochemistry*. 2012;51(45):9178-9191. doi:10.1021/bi300973n.
98. Schlebach JP, Kim M-S, Joh NH, Bowie JU, Park C. Probing membrane protein unfolding with pulse proteolysis. *J Mol Biol*. 2011;406(4):545-551. doi:10.1016/j.jmb.2010.12.018.
99. Minde DP, Maurice MM, Rüdiger SGD. Determining Biophysical Protein Stability in Lysates by a Fast Proteolysis Assay, FASTpp. *PLoS One*. 2012;7(10):1-9.

doi:10.1371/journal.pone.0046147.
100. Na Y, Park C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci*. 2009;18:268-276. doi:10.1002/pro.29.
101. Holtkamp W, Wintermeyer W, Rodnina MV. Synchronous tRNA movements during translocation on the ribosome are orchestrated by elongation factor G and GTP hydrolysis. *Bioessays*. 2014;36:908-918. https://www.pubchase.com/article/25118068.
102. Bhushan S, Meyer H, Starosta AL, et al. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell*. 2010;40(1):138-146. doi:10.1016/j.molcel.2010.09.009.

1.13 Figures

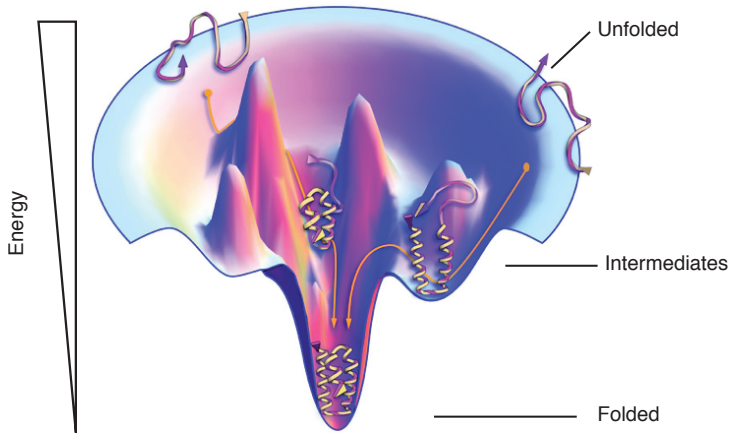


Figure 1.1 – Traditional description of a protein's energy landscape. Most protein energy landscapes are depicted as a funnel. The bottom well contains the native state and, for most proteins, its functional ensemble, which is at the global energy minimum. At the top of the well is the denatured state ensemble. In-between are local minima that sometimes are populated during folding and/or catalysis (adapted from[1]).

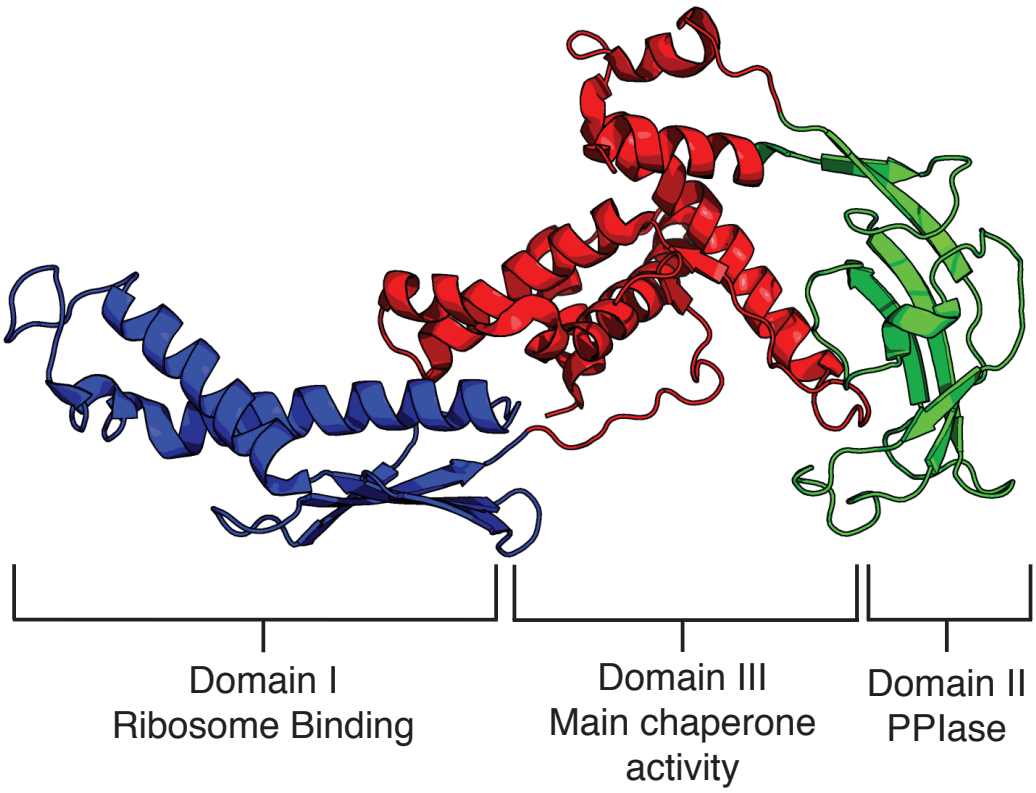


Figure 1.2 - Crystal structure of *V. cholera* Trigger Factor. Trigger factor has three domains: Domain I, which is responsible for ribosome binding and some chaperone activity; Domain II, which is contains the region for trigger factor's main chaperone activity; Domain III, which contains a peptidyl-proline isomerase domain. Structure from PDBID:1T11.

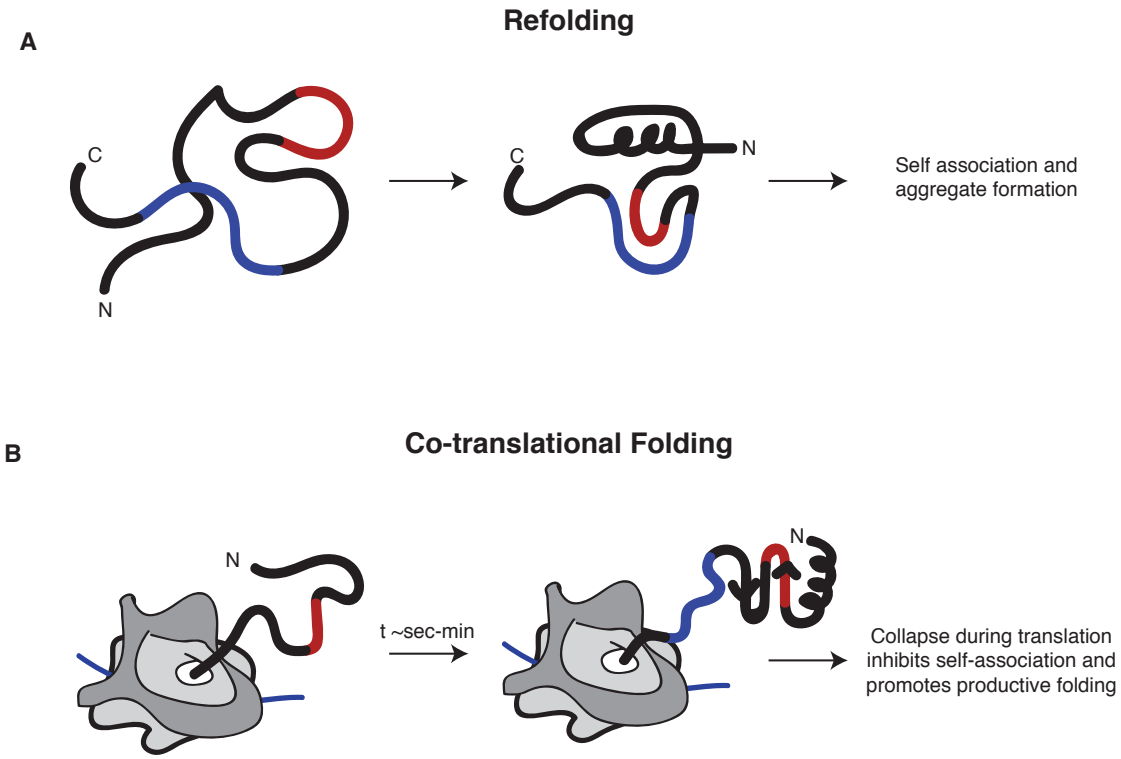


Figure 1.3 – One example scenario of co-translational folding promoting increased folding efficiency. (A) During refolding, two regions of the protein (red and blue) are able to associate and trap the protein in a non-native conformation, possibly leading to aggregation and/or degradation. **(B)** During translation, the red region of the protein has time to collapse and form a stable, on-pathway structure due to the time difference between translation and secondary structure formation. This shields the normally aggregation prone region of the protein and when the blue region emerges, the red region is inhibited from associating with it. Thus, the process of translation can promote folding to the native state.

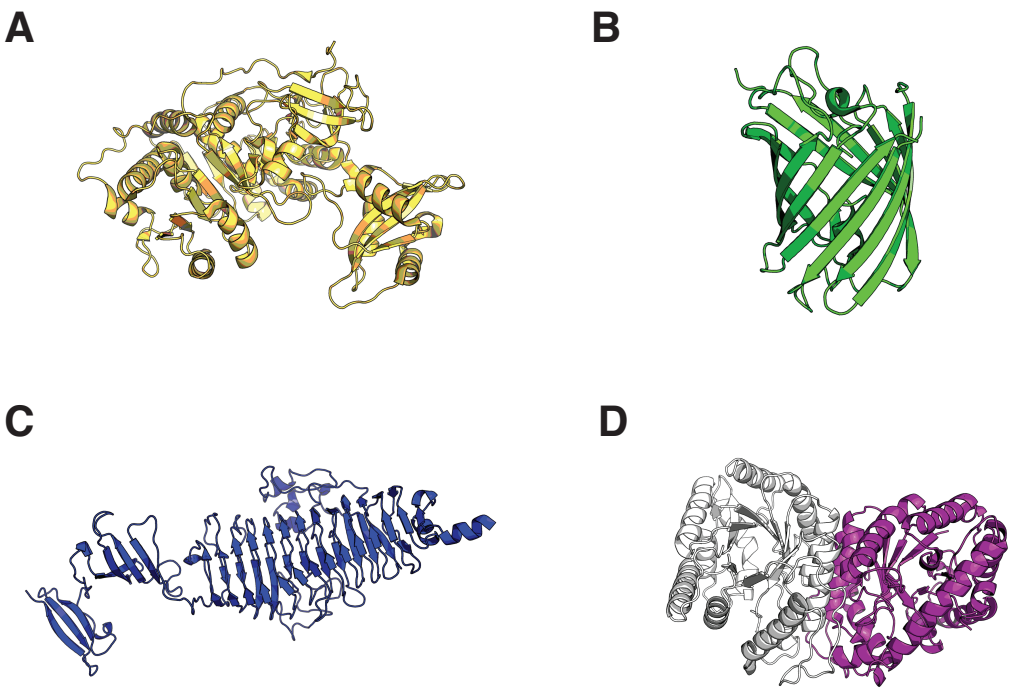


Figure 1.4 Proteins whose folding efficiency has been observed to be changed due to translation – (A) Firefly luciferase (PDBID: 5DV9) **(B)** GFP (PDBID: 2RH7) **(C)** P22-Tailspike (PDBID: 2VFM) **(D)** Bacterial luciferase (PDBID: 1LUC)

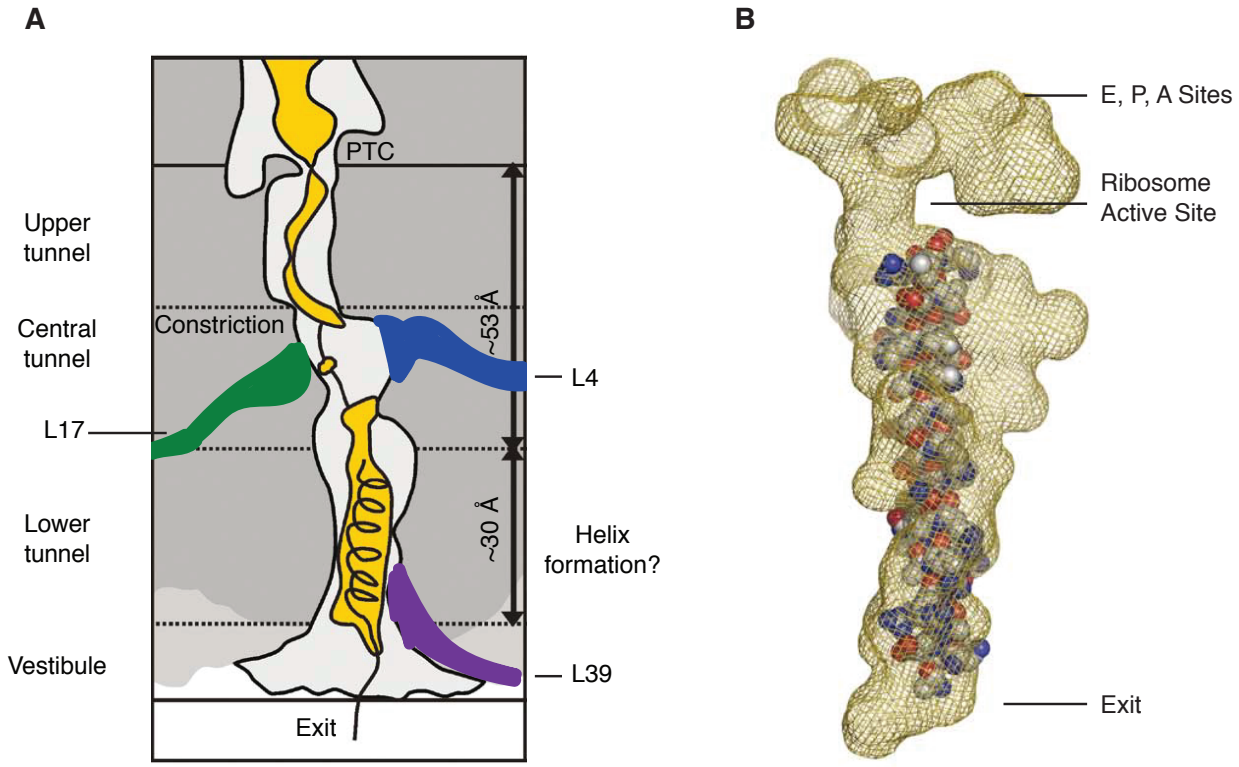


Figure 1.5 The ribosome exit tunnel is big enough to fit an alpha helix – (A) Diagram of ribosome exit tunnel showing proteins L4 (blue), L17 (green) and L39 (purple). The Lower tunnel is large enough to fit an alpha-helix. However, not helix formation can occur prior to the constriction of the ribosome exit tunnel. Figure adapted from Bhushan et al.[102] **(B)** Docking of a model alpha helix within the ribosome exit tunnel. Figure from Voss et al.[69]

Chapter 2
Quantitative determination of ribosome nascent chain stability

2.1 Abstract

Accurate protein folding is essential for proper cellular and organismal function. In the cell, protein folding is carefully regulated; changes in folding homeostasis (proteostasis) can disrupt many cellular processes and have been implicated in a variety of neurodegenerative diseases and other pathologies. For many proteins, the initial folding process begins during translation while the protein is still tethered to the ribosome. Most biophysical studies of a protein's energy landscape, however, are carried out in isolation under idealized, dilute conditions and may not accurately report on the energy landscape under *in vivo* conditions. Thus, the energy landscape of ribosome nascent chains and the effect of the tethered ribosome on nascent-chain folding remain unclear. Here, we develop a general assay for quantitatively measuring the folding stability of ribosome nascent chains and find that the ribosome exerts a destabilizing effect on the polypeptide chain. This destabilization decreases as a function of the distance away from the peptidyl transfer center. Hence, the ribosome may add an additional layer of robustness to the protein-folding process by avoiding the formation of stable partially folded states before the protein has completely emerged from the ribosome.

Note: This is adapted from a paper published in 2016: Samelson AJ et al. *PNAS* (PMID: 27821780)

2.2 Introduction

Proper protein folding is necessary for the function of all cells, and changes in the cellular protein-folding capacity can lead to cell death and disease [1]. For more than 50 years, experimental studies have probed the physical properties of proteins, paving the way for incredible advances in protein design and protein structure prediction, as well as for understanding first principles of protein folding [2,3]. These *in vitro* studies, however, do not necessarily recapitulate the folding process *in vivo,* including the constraints the ribosome imposes on the emerging nascent chain during translation [4].

In the cell, proteins are synthesized by the ribosome one amino acid at a time on a time scale that is slower than most *in vitro* protein folding rates [5]. Thus, the emerging chain has time to explore conformational space and adopt structured conformations before its entire sequence has been synthesized [6]. This vectorial process, and the proximity of the ribosome itself, can modulate the emerging chain's energy landscape in ways that are just beginning to be appreciated: translation can affect folding efficiency, enable the population of intermediates that are not revealed during off-ribosome studies, and can even determine the final conformational fate of the nascent protein [6–10]. In order to understand the folding process *in vivo*, general rules and biophysical mechanisms for these modulations of the emerging chain's energy landscape need to be elucidated.

To unravel the *in vivo* folding landscape, we need to interrogate the energetics and dynamics of ribosome-bound nascent chains in a manner analogous to studies of the *in vitro* folding process. The challenge, however, is that standard methods used to probe protein energy landscapes *in vitro* – such as circular dichroism or intrinsic fluorescence – demand large quantities (milligrams) of extremely pure protein in conditions that differ significantly from those found in the cytosol. Therefore, these experimental approaches are not amenable to buffers that more closely replicate *in vivo* conditions or to heterogeneous protein complexes, such as ribosome nascent chains (RNCs). Recently, approaches such as single-molecule optical trapping and multi-dimensional NMR have revealed interesting effects of the folding of a few different RNCs, but these studies are technically very challenging and therefore are not tractable for characterizing many different proteins [11,12]. Furthermore, these approaches do not provide the most basic features of a protein's energy landscape, such as the global stability of the nascent chain.

Here, we present an approach that we believe takes a first step towards quantitatively measuring the energy landscape of RNCs. Our technique utilizes *in vitro* translation and an easily accessible co-translational labeling scheme in combination with a gel-based method to measure the protein stability of RNCs. We measure the stability of three proteins both on and off the ribosome and demonstrate that the ribosome destabilizes the nascent chain in a manner that is dependent on the distance away from the ribosome's peptidyl-transferase center (PTC). This ribosome-mediated destabilization may safeguard the emerging chain by preventing the formation of stable off-pathway intermediates that lead to misfolding.

2.3 Pulse Proteolysis is a gel-based measurement of protein stability

The simple, first-order description of any protein's energy landscape is its global stability, $\Delta G_{unfolding}$: which determines the relative population of the unfolded and native conformers. This simple parameter is a key determinant for *in vivo* protein lifetimes and many disease-causing SNPs are associated with changes in protein stability [13–15]. *In vitro*, $\Delta G_{unfolding}$ is easily determined by equilibrium denaturation – monitoring the spectroscopic signal from the folded protein as a function of chemical perturbant, such as urea or guanidinium chloride (Figure 2.1). In the presence of the ribosome, however, these simple measurements are not feasible, as the ribosome itself will overwhelm the spectroscopic signal of the nascent chain.

To overcome this, we utilize a gel-based method, pulse proteolysis, which measures protein stability by taking advantage of the fact that unfolded proteins are more susceptible to proteolysis than folded proteins [16]. In pulse proteolysis, an equilibrium mixture of folded and unfolded proteins is subjected to a short pulse of protease, digesting the unfolded proteins and leaving the folded proteins intact. This is then repeated at increasing urea concentrations to alter the equilibrium ratio of folded and unfolded protein. Finally, the fraction of folded protein is determined by SDS-PAGE (Figure 2.1). As long as the protein-unfolding rate is slow compared to the duration of the proteolysis pulse (one minute), the intensity of the full-length band on SDS-PAGE is directly proportional to the equilibrium fraction of native protein prior to digestion. No protease is added to the zero molar urea sample as a control for any proteolysis of folded protein. The intensities of the full length bands at each urea concentration are then fit to determine the denaturant midpoint, or C_m, which when multiplied by a protein's m-value (slope of the unfolding transition) yields the protein's stability. m-Values are not determined by pulse proteolysis and are either calculated (here) or determined using another experimental method [16,17].

2.4 Pulse Proteolysis of *in vitro* translated and labeled proteins.

We performed pulse proteolysis on RNCs by combining a commercially available *in vitro* coupled transcription/translation system (PURExpress NEB, Ipswich, MA) with commercially available BODIPY-FL-Lysine^{AAA}-tRNA (Promega, Madison, WI). This pulse proteolysis/labeling approach can successfully determine the stability of *in vitro* translated proteins that contain a single lysine [18]. To generalize this method, we first compared the stability of proteins containing multiple lysines produced via the same *in vitro* translation strategy or purified recombinantly from *E. coli*. We measured the stabilities of two model proteins by pulse proteolysis: RNase H I53D (eleven lysines) and DHFR V75R (six lysines). Since lysines are predominately surface exposed, we expected BODIPY-FL-lysine incorporation to minimally perturb each protein's native-state stability. We observed a single band for each protein assayed (see top gels Figure 2.2A and 2.2B) and thus assume that only a single BODIPY-lysine is incorporated per protein. Figure 2.2A shows that for DHFR V75R, the stability of the *in vitro* translated, labeled protein was slightly lower than that found for purified recombinant protein. For,

RNase H I53D, the stability of the *in vitro* translated, labeled protein matches that of the unlabeled recombinant protein purified from *E. coli* (Figure 2.2B and Table 1). This slight destabilization may be due to the incorporation of the fluorophore, or from slight differences in the buffer used in the *in vitro* translation reaction compared to those used on the purified protein, as DHFR is very sensitive to changes in salt concentrations. It should be noted, however, that incorporation of BODIPY-FL-lysine does not affect the ability of DHFR V75R to bind one of its inhibitors, methotrexate, suggesting that although the stability of the protein is somewhat decreased, the native conformation of DHFR V75R is not disturbed (See Figure 2.3). While this could pose a problem when comparing purified protein to IVT translated protein, it should not affect a direct comparison of IVT-produced protein on- and off- the ribosome.

2.5 Urea sensitivity of 70S ribosome and RNCs.

Determining protein stability (ΔG_{unf}) requires modulating and measuring the population of both the folded and unfolded states; this is usually done using thermal or chemical denaturation. Urea is the preferred denaturant for pulse proteolysis, because it does not alter the samples' ionic strength and can be used in combination with SDS-PAGE. Considering its ubiquity in protein-folding experiments, it is surprising that, to our knowledge, no group has determined the sensitivity of RNCs to urea denaturation. To probe this, we used both sucrose gradient ultracentrifugation and fluorescence correlation spectroscopy (FCS).

First, we performed sucrose gradient ultracentrifugation on purified 70S ribosomes after incubation in varying urea concentrations (Figure 2.4A). 70S ribosomes appear stable in 1.0M urea, although the peak shifts down the gradient marginally, perhaps signifying slight expansion or partial unfolding of ribosomal proteins or rRNA. At 2.0M urea, however, the 70S peak decreases concurrent with an increase in the heights of both the 30S and 50S peaks. The 70S peak then completely disappears at 3.0M urea and the 50S peak broadens and simultaneously decreases in intensity, signifying denaturation of 50S subunits.

SecM-stalled ribosomes should be more stable than reconstituted 70S ribosomes since the SecM stalling sequence is known to make both extensive contacts with the ribosome exit tunnel and strengthen 30S-50S contacts [19]. To test the urea sensitivity of SecM-stalled RNCs, we translated DHFR V75R-$(GS)_5$-SecM for one hour, halted translation with 2mM chloramphenicol, incubated the IVT reactions in urea overnight, and subjected these samples to sucrose-gradient ultracentrifugation. At 0M and 1.0M urea, as observed for reconstituted 70S ribosomes, the height of the 70S peak changes little for SecM-stalled RNCs. Unlike reconstituted 70S ribosomes, however, a significant portion of the 70S peak remains for SecM-stalled RNCs incubated at 2.0M and 3.0M urea (compare Figure 2.4A and 2.4B). It is possible that a small population of the 70S ribosomes is not properly engaged with SecM and this could account for the small decrease in 70S peak intensity between 1.0M and 2.0M urea. At 4.0M urea, there is no apparent 70S peak. In sum, these data suggest that the SecM-stalled RNCs are stable

to greater than 3.0M urea. To confirm this observation, we turned to FCS in order to follow RNCs specifically, rather than the entire 70S ribosome population.

FCS studies were carried out on RNCs labeled with Atto647N via unnatural amino acid incorporation (Figure 2.4C) and were incubated overnight in urea as described above. From 0M urea to 3.0M urea, the small increase in the measured diffusion coefficient is likely due to a small population of free, dye-labeled nascent chains or a small proportion of RNCs improperly engaged with SecM, as is seen in the previously mentioned sucrose gradient experiments. Since we use a single-species fit (see Methods), a small proportion of free nascent chains would appear as an increase in the observed diffusion coefficient. The diffusion coefficient measured in low concentrations of urea is about 2.0 $x10^{-10}$ m^2sec^{-1}, close to the diffusion coefficient measured for 70S ribosomes using laser light-scattering microscopy [20], although in different conditions. Above 3.0M urea, there is a dramatic increase in the observed diffusion coefficient - likely a result of 70S dissociation into 30S and 50S subunits. At 4.0M urea, the diffusion coefficient measured corresponds to that of sphere with mass ~40kD (approximately the molecular weight of the protein and attached tRNA) and remains smaller than that of nascent chains released by RNase A digestion. These data support those seen in the sucrose gradient ultracentrifugation experiments – SecM stalled RNCs are stable to greater than 3.0M but less than 4.0M urea.

2.6 Pulse proteolysis of stalled ribosome nascent chains.

In order to determine if the ribosome can alter the stability of the emerging chain, we applied pulse proteolysis to stalled RNCs. We used protein variants with C_ms lower than 2.5M urea off the ribosome to assure that the unfolding transition occurs before RNC dissociation: RNase H I53D, DHFR V75R and Barnase W35F/W94F/H102A. All three of these variants have been well characterized *in vitro* and display two-state equilibrium unfolding [21,22]. For each protein tested, RNCs are destabilized as compared to the same protein free from the ribosome (Figure 2.5 A-C, Table 2). In addition, the magnitude of destabilization is anti-correlated to the isoelectric point of the protein. DHFR V75R is destabilized by 1.93±0.29 kcal/mol and has an isoelectric point of 4.94, whereas RNase H I53D and Barnase W35F/W94F/H102M, whose isoelectric points are 7.97 and 8.86, are destabilized by 1.63±0.36 kcal/mol and 0.42±0.08 kcal/mol, respectively. These data are consistent with previous studies on ribosome-stalled intrinsically disordered proteins (IDPs) which have demonstrated that more negatively charged IDPs have increased dynamic motions when stalled on the ribosome as compared to when free in solution [23]. Similarly, NMR data show an increase in dynamics of the nascent chain when proximal to the ribosome [12,24]. Both of these results can be explained by a general destabilization of the nascent chain due to proximity to the ribosome.

To assure that our proteolysis measurements are not biased due to steric occlusion of the protease by the ribosome, we appended a TEV-cleavage site between the natural C-terminus of the protein and the glycine-serine linker. We monitored proteolysis kinetics using TEV protease on RNCs both before and after RNaseA

treatment. In both cases, the protein is fully cleaved within 5 minutes of TEV addition (Figure 2.6). Thus, even the C-termini of the RNCs are fully accessible to proteolysis.

2.7 Reversible folding of stalled nascent chains.

Measuring thermodynamic stability requires that the folding process is reversible and that the proteolysis-resistant conformation does not represent a kinetic trap. To test for reversibility, after purification through a sucrose cushion, we re-suspended stalled DHFR V75R-$(GS)_5$-SecM RNCs in both 0.5M and 2.5M urea and allowed them to reach equilibrium overnight. Each RNC/urea sample was then divided and diluted into either 2.5M urea or 0.5M urea (see Figure 2.7). After a second overnight equilibration, the fraction of folded protein was measured by pulse proteolysis. Regardless of the initial urea concentration, the fraction folded of RNC at 0.5M urea remained the same whether the sample had been unfolded in 2.5M urea or not (Figure 2.7 A and C). Therefore, stalled nascent chains fold reversibly.

2.8 Nascent chain destabilization is distance dependent.

Several studies have suggested a distance dependence to the effects of ribosome-mediated changes in nascent chain behavior [6,11,12]. To investigate the distance dependence of the stability of RNCs, we used both DHFR V75R and RNase H I53D to measure protein stability as a function of amino acid distance from the peptidyl-transferase center (PTC) by increasing the linker length between the SecM stalling sequence and the natural C-terminus of the protein (See Figure 2.8E). The linker length was extended by increasing the number of glycine-serine repeats in steps of ten amino acids, resulting in distances of 35, 45 and 55 amino acids between the natural C-terminus of the target protein and the PTC. For both proteins, nascent chain stability increased as a function of linker length, approaching the stability of the free protein at a distance of 55 residues away from the PTC (Figure 2.8 A-D, Table 3). These results can explain previous observations of both increased protection to limited proteolysis and increased peak dispersion as the distance from the PTC is increased [12,25–27], suggesting that the increased protection is likely due to changes in global stability and not interactions with the ribosome or changes in native-state dynamics.

2.9 Discussion

Here, we developed a simple gel-based assay for measuring the stability of RNCs and compared the global stability of three proteins both on the ribosome and free in solution. We provide what we believe are the first quantitative measurements of a protein's global stability on the ribosome and find that RNCs are destabilized relative to the same protein off the ribosome. This ribosome-dependent modulation of the energy landscape is dependent on the amino acid distance from the peptidyl-transferase

center. By the time the nascent chain is 55 residues away from the PTC (a spacer of 20-30 residues from the end of the exit tunnel), the global stability of the protein is no longer modulated by the presence of the ribosome. These results are consistent with, and provide a quantitative explanation for, several other observations of ribosome nascent chain behavior [6,11,12,28]. For instance, the point at which the nascent chain acquires its native stability as measured here is very similar to the distance from the PTC required for full acquisition of folded peaks by NMR [12]. Furthermore, work on stalled IDPs suggests that the ribosome acts comparably to a denaturant on the emerging chain, which again is consistent with what we observe here for the folding energetics of RNCs [23].

It is particularly interesting to compare our results with those obtained by von Heijne and colleagues using an arrest peptide mediated assay to assess the force generated by the nascent chain during translation [28,29]. For DHFR, they measure the maximum force generated by the emerging chain at about 45 amino acids away from the PTC, close to where we see a return to off-ribosome like stability. It remains unclear, however, how the force applied to the arrest peptide is related to the protein's stability, folding trajectory, or folding rates. Our approach, which can be used under various conditions and at a range of nascent-chain lengths, should help shed light on the biophysical effects driving the peptide arrest assay.

Several other studies have implicated electrostatic forces in modulating nascent chain energy landscapes [11,12,23]. The distance dependence of the stability changes we observe, combined with the fact that the magnitude of destabilization is inversely correlated with the isoelectric point of the proteins we measure, are consistent with this idea. Further studies, however, are needed to determine the physical basis of the destabilization observed here.

What role could general ribosomal destabilization play in assuring folding fidelity *in vivo*? For the emerging chain, accessing the native state is not an option, simply because the full sequence of the protein is yet to be synthesized. However, hydrophobic collapse or other partially folded states are accessible to the incomplete nascent chain, and translation is certainly slow enough that the incomplete nascent chain has time for formation of such potential native and non-native intermediates. While some experiments have shown co-translational folding to increase folding efficiency and speed, it is possible that intermediates that form co-translationally may be off-pathway and result in non-native, toxic species [7,30,31]. In order to avoid such conformations, cells may have evolved chaperones, such as trigger factor, to "hold and unfold" proteins as they emerge. Indeed, trigger factor is known to bind to proteins about 60 amino acids away from the PTC, the same distance at which the destabilization we observe is abrogated [26,27]. Perhaps the observed destabilization allows the nascent chain to fold more efficiently and avoid off-pathway kinetic traps. Therefore, destabilizing collapsed states that form as the nascent peptide emerges, but before its entire sequence is accessible, may help to ensure that the protein only forms a stable structure after completely emerging from the ribosome.

Finally, our approach is adaptable as a tool to probe other differences between protein energy landscapes *in vitro* and *in vivo*. While we believe this method to be widely applicable; when applying it to new systems, it is important to take into account

the limitations of pulse proteolysis, including possible steric hindrance due to biomolecular interactions (see 16, 32). Nevertheless, it should be possible to extend our approach to determine how the ribosome modulates protein folding kinetics by monitoring unfolding rates via pulse proteolysis [33]. In addition, since quality control at the ribosome seems to play an important role in general cellular proteostasis [34], it will be important to know how RNC stability is altered in the presence of ribosome-associated chaperones such as trigger factor (in bacteria) or the RQC (in eukaryotes). Importantly, moving from more descriptive studies of RNCs to measuring their biophysical properties will enable both validation and extension of recent computational studies [35–37] and will uncover general biophysical principles governing co-translational folding.

2.10 Materials and Methods

Sample preparation 50μL IVT reactions (PURExpress, NEB) were initiated by addition of 500ng plasmid DNA in the presence of 2.5μL Flourotect Greenlys tRNA (Promega) and 2μL RNase inhibitor, murine (NEB).

Preparation of samples off the ribosome or without a SecM sequence:
Samples were incubated for 1 hr at 37°C before adding chloramphenicol to 2mM and RNase A to a final concentration of 0.1 mg/mL. These samples were then incubated at room temperature overnight and spun at 21000xg for 30 minutes at 4°C. The supernatant was then used for pulse proteolysis.

Preparation of RNCs:
After incubation for 30 minutes to 1 hour at 37°C, IVT reactions were loaded onto a 125μL 1M sucrose cushion in 25mM HEPES pH 7.5, 15mM MgOAc, 150mM KCl, 2mM DTT (HKM+DTT) and centrifuged at 200,000xg for 40 minutes at 4°C. Supernatant was aspirated and ribosome pellets were washed three times with 200μL HKM+DTT. Pellets were then resuspended in 35μL of HKM +DTT (see Figure 2.9).

Pulse Proteolysis
For protein purified from *E. coli*, pulse proteolysis was conducted as previously described [16,32] in HKM+DTT. For released or stalled nascent chains, 3μL of halted IVT reactions or RNCs, respectively, were diluted into 7μL of HKM+DTT and urea to the desired urea concentration. After incubation for at least 12 hours, 1μL of 34x (6.8mg/mL) thermolysin was added to each 10μL reaction and 8μL were quenched into 3μL 500 mM EDTA pH 8.5. After pulse proteolysis, RNase A was added to 0.1mg/mL to each reaction and incubated at 37°C overnight to digest any remaining peptidyl-tRNA. For IVT reactions off the ribosome, RNase A was added to a final concentration of 0.1mg/mL and incubated 15 minutes at 37°C. Samples were then mixed with SDS-PAGE loading dye and loaded onto 4-12% Bis-Tris gels (ThermoFisher, Waltham, MA). Gels were run in MES buffer and imaged with a typhoon (GE, Schenectady, NY) using a 488nm laser and 520BP filter. Analysis and quantification of gels was done using

ImageJ as described previously [32]. Urea concentrations were measured using a refractometer as described previously [32].

FCS

RNCs with fluorescently labeled nascent chains were a gift of Madeleine Jensen. For experiments, they were diluted into appropriate urea concentrations, and allowed to reach equilibrium overnight at room temperature in 1x HKM+DTT. FCS measurements and analysis were performed as described previously [38] fitting to a single species, using an additional term to correct for the triplet state. To control for effects of urea on optics and viscosity, diffusion of free Alexa 488 was measured at the same urea concentrations as the RNCs (See Figure 2.10). The measured Alexa 488 diffusion coefficients were then normalized to the 0M urea coefficient to determine the viscosity. These values were used to calculate RNC diffusion coefficients.

Sucrose Gradients

70S ribosomes with no nascent chain were a gift of Jonas Noeske and prepared as in [39]. For RNCs, IVT reactions were quenched with chloramphenicol to a final concentration of 2mM after preparation as described above, diluted to 100μL to the desired urea concentrations, and incubated for at least 12 hours at room temperature. Samples were then layered on top of 10-50% sucrose gradients and centrifuged in a SW41Ti rotor at 40000 rpm for 3 hours.

2.11 Acknowledgements

We would like to thank K. Hamadani and the entire Marqusee lab for advice throughout the project. We would like to thank J.S. Fraser for comments on the manuscript and D. Weiner and C. Reidel for help with FCS data collection and analysis.

2.12 References

1. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426(6968):884-890. doi:10.1038/nature02261.
2. Clark PL. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci*. 2004;29(10):527-534. doi:10.1016/j.tibs.2004.08.008.
3. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591.
4. Braselmann E, Chaney JL, Clark PL. Folding the proteome. *Trends Biochem Sci*. 2013;38(7):337-344. doi:10.1016/j.tibs.2013.05.001.
5. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. *Curr Opin Struct Biol*. 2010;20(1):33-45. doi:10.1016/j.sbi.2010.01.005.
6. Holtkamp W, Kokic G, Jäger M, Mittelstaet J, Komar AA, Rodnina M V. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104-1107. doi:10.1126/science.aad0344.
7. Ugrinov KG, Clark PL. Cotranslational folding increases GFP folding yield. *Biophys J*. 2010;98(7):1312-1320. doi:10.1016/j.bpj.2009.12.4291.

8. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858-861. doi:10.1021/ja411302m.
9. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009;16(3):274-280. doi:10.1038/nsmb.1554.
10. Hsu S-TD, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2007;104(42):16516-16521. doi:10.1073/pnas.0704664104.
11. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C. The ribosome modulates nascent protein folding. *Science*. 2011;334(6063):1723-1727. doi:10.1126/science.1209740.
12. Cabrita LD, Cassaignau AME, Launay HMM, et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol*. 2016;(February). doi:10.1038/nsmb.3182.
13. Wang Z, Moult J. SNPs, protein structure, and disease. *Hum Mutat*. 2001;17(4):263-270. doi:10.1002/humu.22.
14. Yue P, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*. 2005;353(2):459-473. doi:10.1016/j.jmb.2005.08.020.
15. Parsell DA, Sauer RT. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J Biol Chem*. 1989;264(13):7590-7595. http://www.ncbi.nlm.nih.gov/pubmed/2651442.
16. Park C, Marqusee S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*. 2005;2(3):207-212. doi:10.1038/nmeth740.
17. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci*. 1995;4(10):2138-2148. doi:10.1002/pro.5560041020.
18. Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol*. 2012;8(2):147-153. doi:10.1038/nchembio.742.
19. Mitra K, Schaffitzel C, Fabiola F, Chapman MS, Ban N, Frank J. Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol Cell*. 2006;22(4):533-543. doi:10.1016/j.molcel.2006.05.003.
20. Gabler R, Westhead EW, Ford NC. Studies of ribosomal diffusion coefficients using laser light-scattering spectroscopy. *Biophys J*. 1974;14(7):528-545. doi:10.1016/S0006-3495(74)85933-3.
21. Garvey EP, Matthews CR. Effects of multiple replacements at a single position on the folding and stability of dihydrofolate reductase from Escherichia coli. *Biochemistry*. 1989;28(5):2083-2093. http://www.ncbi.nlm.nih.gov/pubmed/2655702.
22. Spudich GM, Miller EJ, Marqusee S. Destabilization of the Escherichia coli RNase

H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol*. 2004;335(2):609-618. doi:10.1016/j.jmb.2003.10.052.
23. Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, Cavagnero S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem Biol*. 2013;8(6):1195-1204. doi:10.1021/cb400030n.
24. Eichmann C, Preissler S, Riek R, Deuerling E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2010;107(20):9111-9116. doi:10.1073/pnas.0914300107.
25. Tomic S, Johnson AE, Hartl FU, Etchells SA. Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. *FEBS Lett*. 2006;580(1):72-76. doi:10.1016/j.febslet.2005.11.050.
26. Hoffmann A, Merz F, Rutkowska A, Zachmann-Brand B, Deuerling E, Bukau B. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J Biol Chem*. 2006;281(10):6539-6545. doi:10.1074/jbc.M512345200.
27. Hoffmann A, Becker AH, Zachmann-Brand B, Deuerling E, Bukau B, Kramer G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol Cell*. 2012;48(1):63-74. doi:10.1016/j.molcel.2012.07.018.
28. Nilsson OB, Hedman R, Marino J, et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep*. 2015;12(10):1533-1540. doi:10.1016/j.celrep.2015.07.065.
29. Nilsson OB, Müller-Lucks A, Kramer G, Bukau B, von Heijne G. Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein. *J Mol Biol*. 2016;428(6):1356-1364. doi:10.1016/j.jmb.2016.02.014.
30. Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*. 1999;6(7):697-705. doi:10.1038/10754.
31. Deckert A, Waudby CA, Wlodarski T, et al. Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. *Proc Natl Acad Sci U S A*. 2016;113(18):5012-5017. doi:10.1073/pnas.1519124113.
32. Park C, Marqusee S. Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr Protoc Protein Sci*. 2006;Chapter 20:Unit 20.11. doi:10.1002/0471140864.ps2011s46.
33. Na Y-R, Park C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci*. 2009;18(2):268-276. doi:10.1002/pro.29.
34. Choe Y-J, Park S-H, Hassemer T, et al. Failure of RQC machinery causes protein aggregation and proteotoxic stress. *Nature*. 2016;531(7593):191-195. doi:10.1038/nature16973.
35. O'Brien EP, Vendruscolo M, Dobson CM. Prediction of variable translation rate effects on cotranslational protein folding. *Nat Commun*. 2012;3(May):868. doi:10.1038/ncomms1850.
36. Sharma AK, Bukau B, O'Brien EP. Physical Origins of Codon Positions That Strongly Influence Cotranslational Folding: A Framework for Controlling Nascent-

Protein Folding. *J Am Chem Soc*. 2016;138(4):1180-1195. doi:10.1021/jacs.5b08145.
37. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc*. 2011;133(3):513-526. doi:10.1021/ja107863z.
38. Riedel C, Gabizon R, Wilson C a M, et al. The heat released during catalytic turnover enhances the diffusion of an enzyme. *Nature*. December 2014. doi:10.1038/nature14043.
39. Noeske J, Wasserman MR, Terry DS, Altman RB, Blanchard SC, Cate JHD. High-resolution structure of the Escherichia coli ribosome. *Nat Struct Mol Biol*. 2015;22(4):336-341. doi:10.1038/nsmb.2994.

2.13 Figures

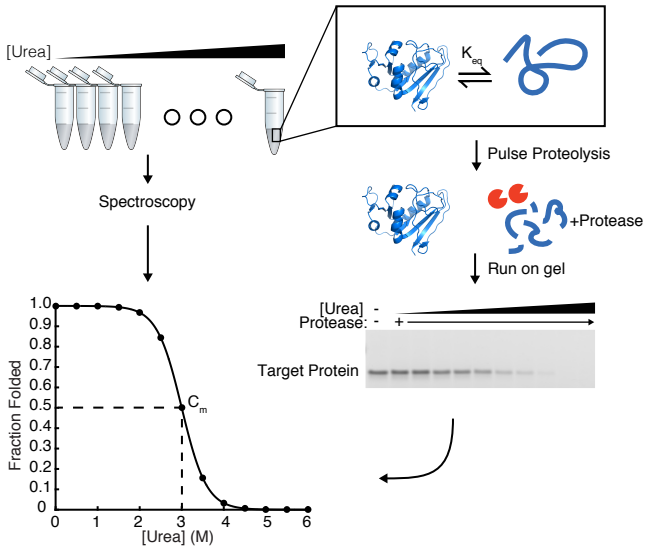


Figure 2.1 - Comparison of spectroscopy and pulse proteolysis for determining protein stability. Samples are equilibrated in urea overnight for both spectroscopic methods and pulse proteolysis For spectroscopic methods (left), equilibrated samples are read directly in a spectrophotometer to determine the signal difference between the folded and unfolded states. Alternatively, for pulse proteolysis (right), a protease is added to digest all unfolded protein at a given urea concentration. Samples are analyzed by SDS-PAGE, and the remaining band represents the amount of folded protein at each urea concentration. The intensity of each band is plotted against the urea concentration and then fit to determine the concentration of urea where half of the folded protein remains, or the C_m. The C_m is multiplied by the protein's m-value to determine the stability, $\Delta G_{unfolding}$.

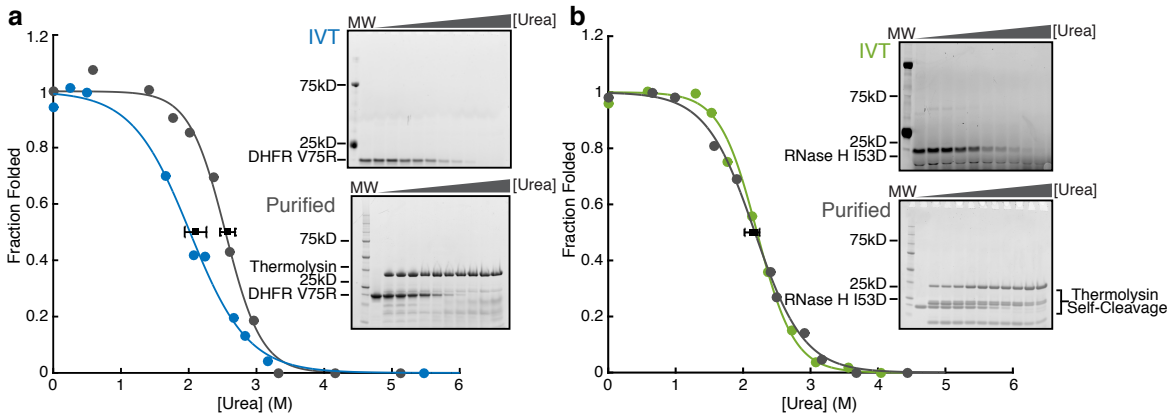


Figure 2.2 - Stability of proteins purified from *E. coli* and made using *in vitro* translation by pulse proteolysis. (a) DHFR V75R purified (in grey and upper gel) and *in vitro* translated (in blue and lower gel). **(b)** RNase H I53D purified (in grey and upper gel) and *in vitro* translated (in green and lower gel). Note that concentration of urea is different for each lane between gels. Each gel and trace shown is representative of three separate experiments. Error bars represent the standard deviation (SD) of the C_m for each curve plotted, determined by three separate experiments.

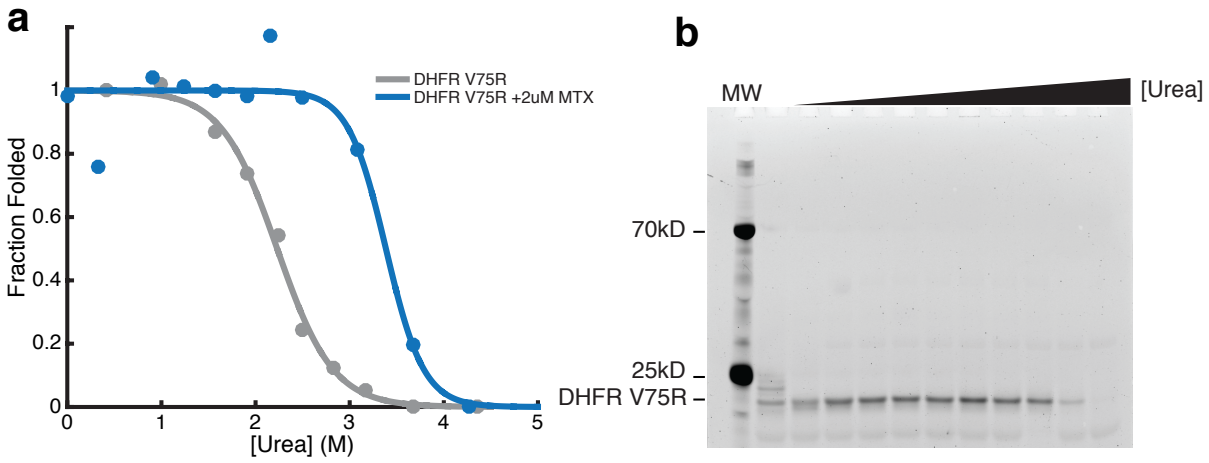


Figure 2.3 - Methotrexate (MTX) binds *in vitro* translated DHFR V75R. (a) Grey, fraction folded as a function of urea for *in vitro* translated DHFR V75R. Blue, fraction folded as a function of urea for in vitro translated DHFR V75R in the presence of 2μM MTX. **(b)** Gel used for quantitation of +MTX sample in **(a)**.

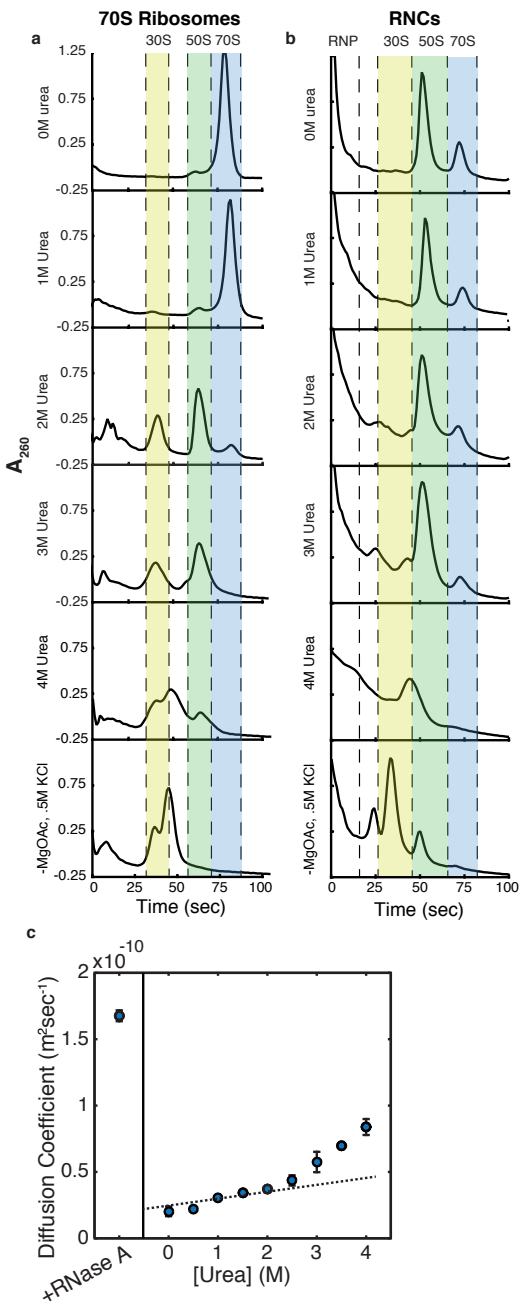


Figure 2.4 - Urea sensitivity of 70S ribosomes and RNCs. (a and b) Sucrose gradient ultra-centrifugation of **(a)** 70S ribosomes and **(b)** RNCs. Highlighted in blue is the peak corresponding to 70S ribosomes. The 50S peak is highlighted in green and the 30S peak in yellow. 70S ribosomes and RNCs were also run in gradients containing no magnesium and 0.5M KCl as a negative control. **(c)** Diffusion coefficients of RNCs plotted as a function of urea concentration (bottom). Error bars represent the standard deviation of at least ten experiments (fits are shown in Figure 2.11).

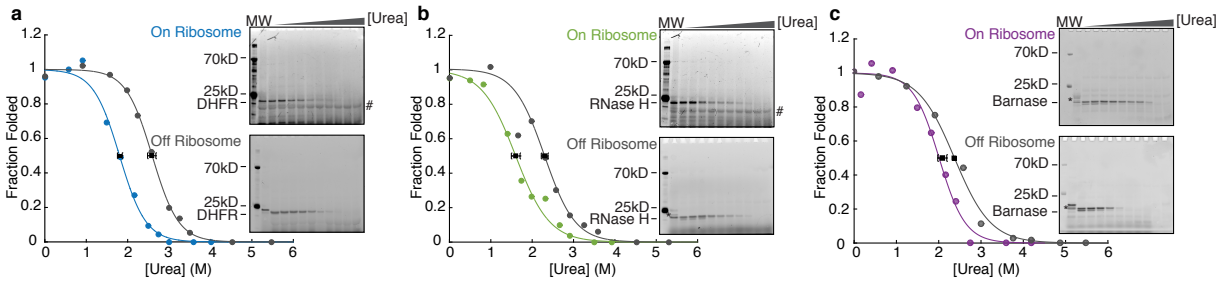


Figure 2.5 - Determination of RNC stability by pulse proteolysis. (a) DHFR V75R RNCs. Blue is protein on the ribosome; Grey is protein off the ribosome. **(b)** RNase H I53D RNCs. Green, on the ribosome; grey, off the ribosome. **(c)** Barnase W35F/W94F/H102M RNCs. Purple, on the ribosome; grey, off the ribosome. Each gel and trace shown is representative of three separate experiments. Asterisk (*) is incomplete tRNA digestion. Hash (#) is RNase A. Error bars represent the standard deviation of the C_m for each curve determined by three separate experiments, except for Barnase, where the C_m is the standard deviation of two experiments.

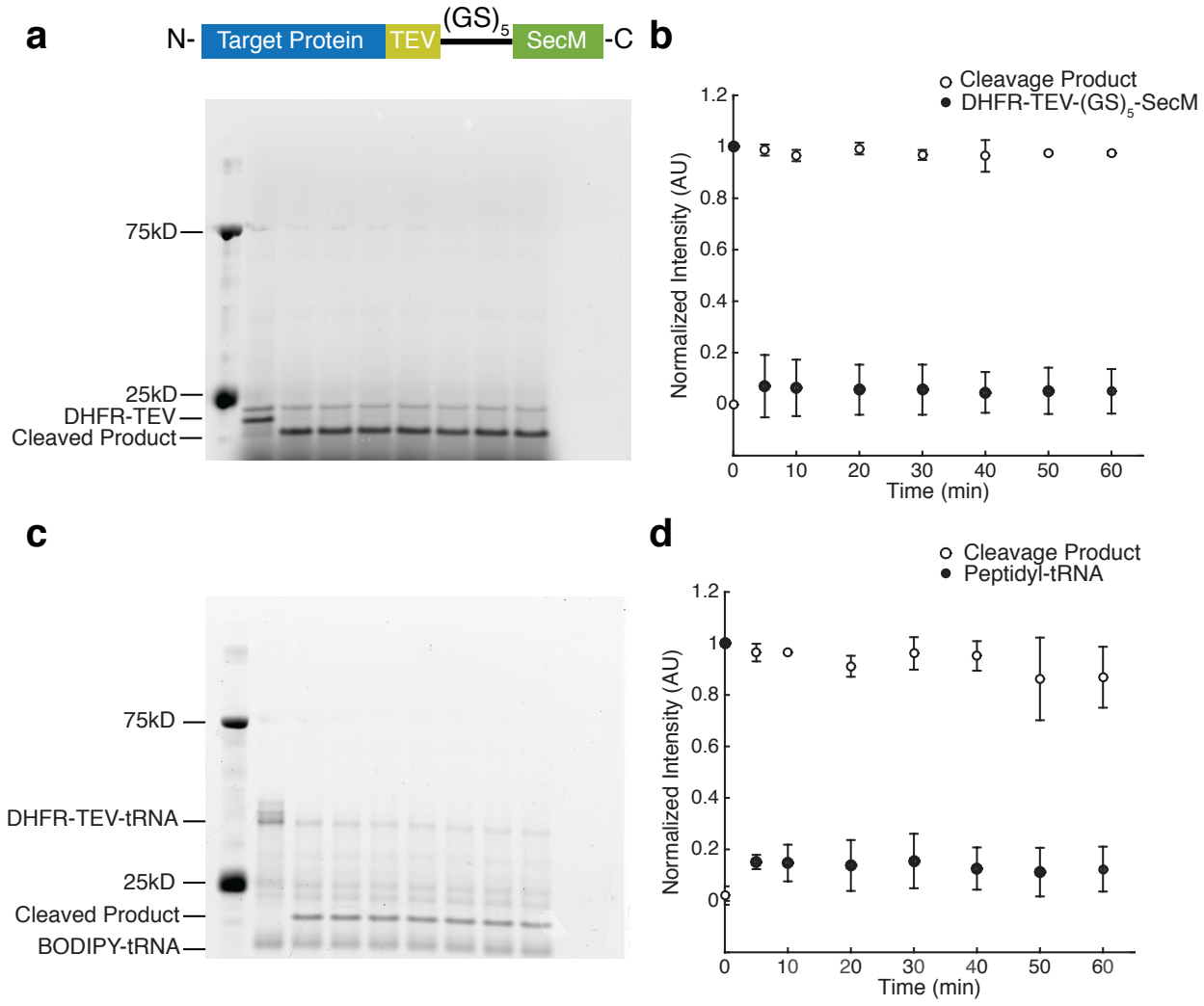


Figure 2.6 - The presence of the ribosome does not inhibit protease accessibility. **(a)** TEV cleavage of DHFR V75R-TEV-(GS)5-SecM as a function of time after RNase A digestion overnight. **(b)** Quantification of gel band intensities normalized to the highest band intensity on the gel. Original band is shown as filled circles and the band corresponding to cleavage product is show as open circles. **(c)** and **(d)** same as **(a)** and **(b)** but as RNCs. Each gel and point shown is representative of three separate experiments. Error bars represent the standard deviation of the normalized band intensity for each time point.

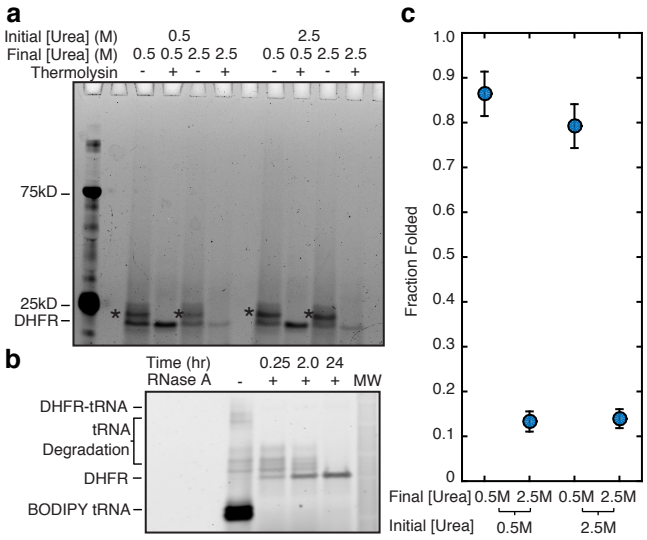


Figure 2.7 - Folding on the ribosome is reversible. DHFR V75R RNCs reached equilibrium in either 0.5M or 2.5M urea. Samples were then split in half and diluted to 0.5M urea or 2.5M urea. After another equilibration step, each sample was again split in half and either treated with or without thermolysin to assess the amount of folded protein remaining and run on a gel **(a)**. Starred bands are due to incomplete RNase A digestions of attached tRNA. **(b)** Gel showing complete cleavage of peptidyl-tRNA as a function of time. **(c)** Quantitation of data shown in (a). Since there is the same amount of folded protein independent of the initial concentration of urea, folding is reversible. Error bars represent the standard deviation of duplicate experiments.

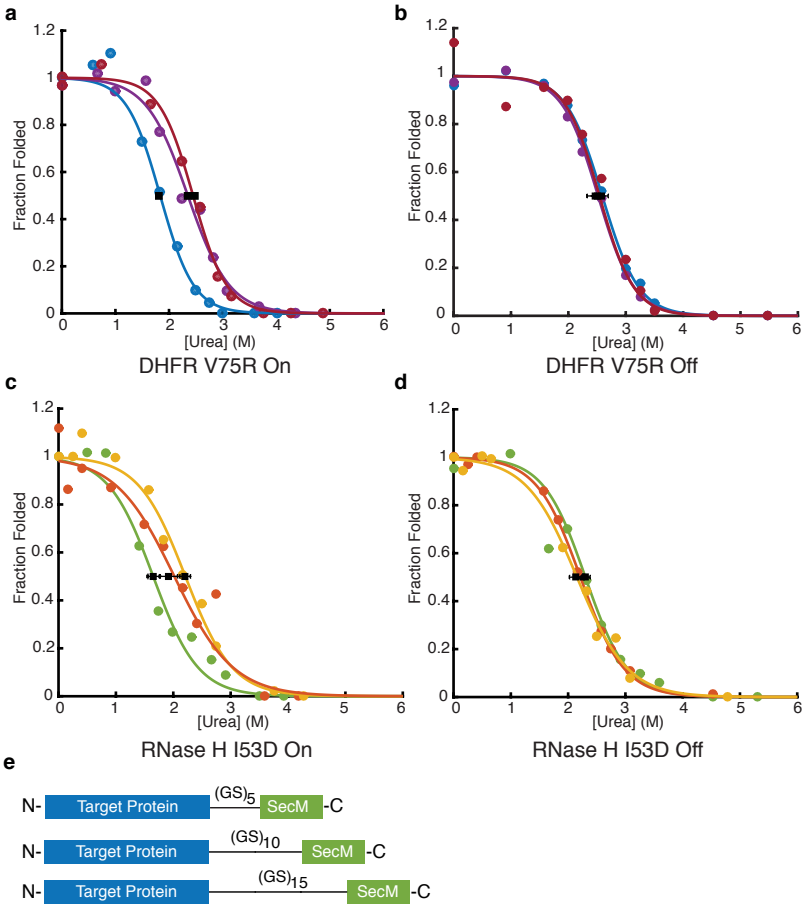


Figure 2.8 - RNC stability increases as the distance to the PTC increases as determined by pulse proteolysis. (a) DHFR V75R RNC stability as a function of linker length. Blue, 35 amino acids from PTC; purple, 45 amino acids; red, 55 amino acids. **(b)** Same as (a) but with a stalling deficient SecM mutant: ^{1}FSTPVWISQAQGIAAGA^{17}. **(c)** RNase H I53D RNC stability as a function of linker length. Green, 35 amino acids from PTC; orange, 45 amino acids; yellow 55 amino acids. **(d)** Same as (c) except after RNase A digestion overnight. Each trace shown is representative of three separate experiments. Error bars represent the standard deviation of the C_m determined by three separate experiments. **(e)** Constructs used in this study. The sequence coding for each target protein was appended with a variable size glycine-serine linker, $(GS)_x$, followed by the SecM stalling sequence at its C-terminus. Gels used are in Figure 2.12

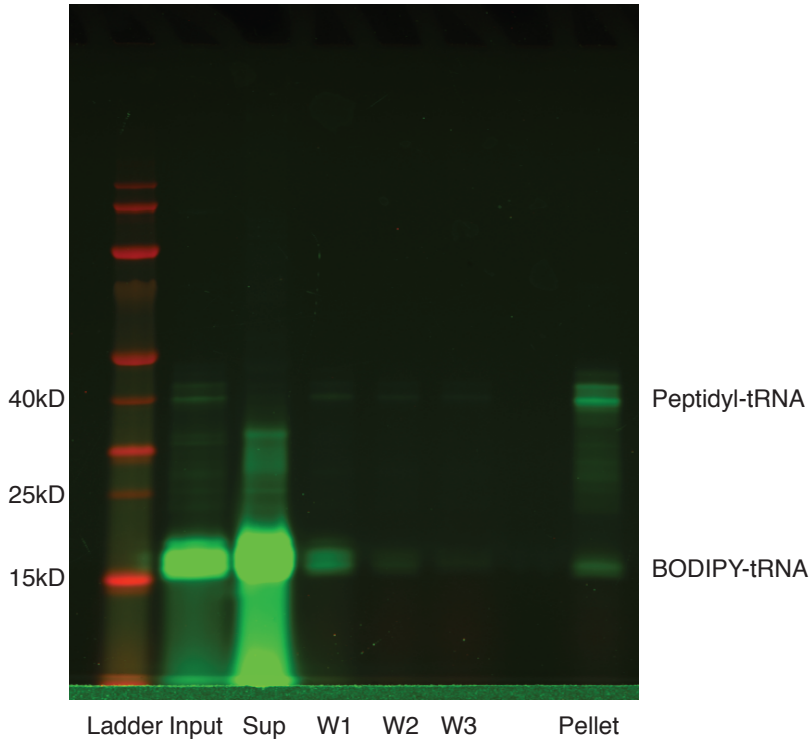


Figure 2.9 - **Purification of labelled RNCs.** IVT reactions (input) were loaded onto a sucrose cushion and centrifuged as described in the methods section. Supernatant (sup) was aspirated and the pellets were washed 3 times (W1, W2, W3) with 200µL of ice cold HKM +DTT. Pellets were then resuspended, mixed with SDS-PAGE loading dye and loaded on a gel.

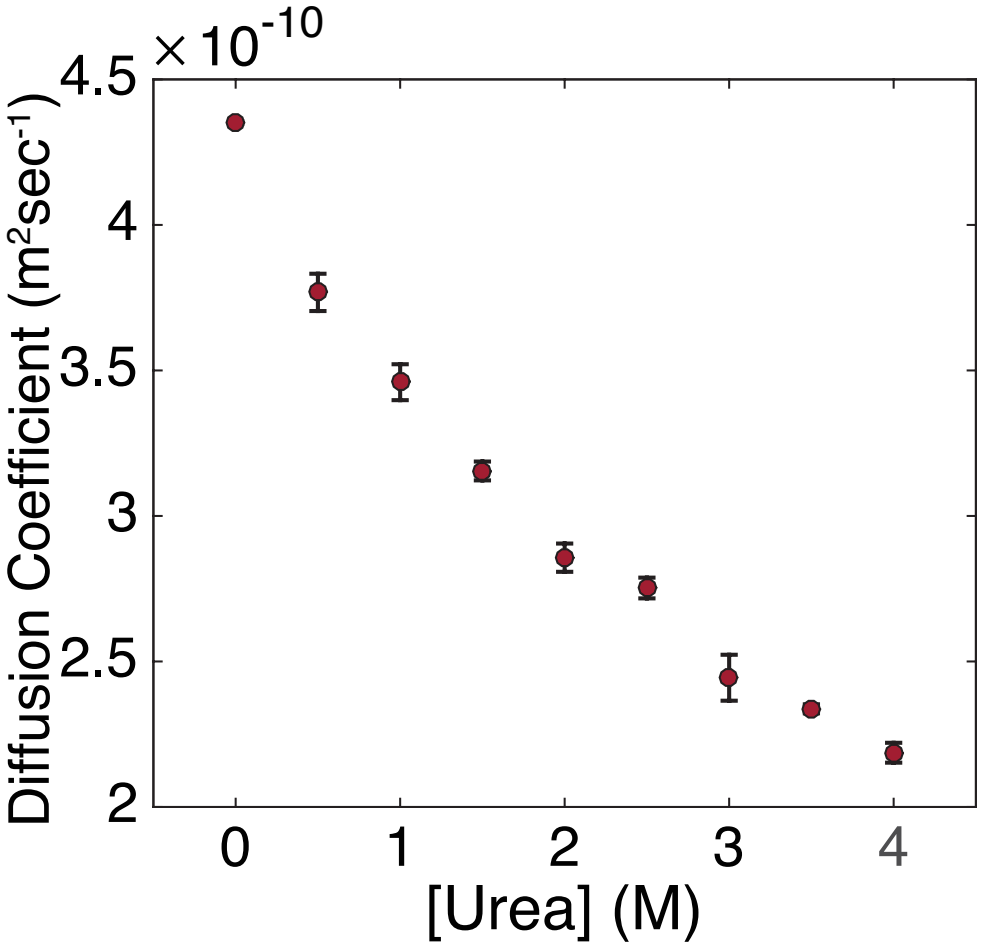


Figure 2.10 - Diffusion of Alexa488 as a function of urea concentration. The diffusion of Alexa488 was used to determine the viscosity of the solution when calculating diffusion coefficients of RNCs. Shown here is the apparent diffusion coefficient without correcting for changes in viscosity due to the presence of high concentrations of urea. Error bars are standard deviations of 15 experiments.

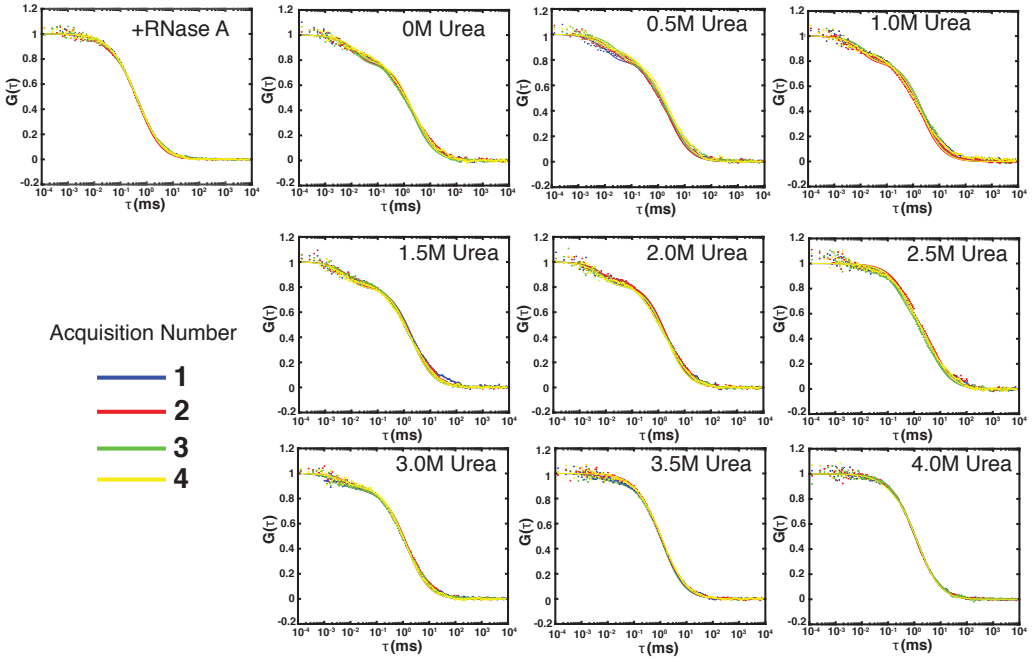


Figure 2.11 - Sample FCS runs with fits for each urea concentration. Colors are four different acquisitions. Diffusion coefficients of RNCs plotted as a function of urea concentration (bottom). Error bars represent the standard deviation of at least ten experiments.

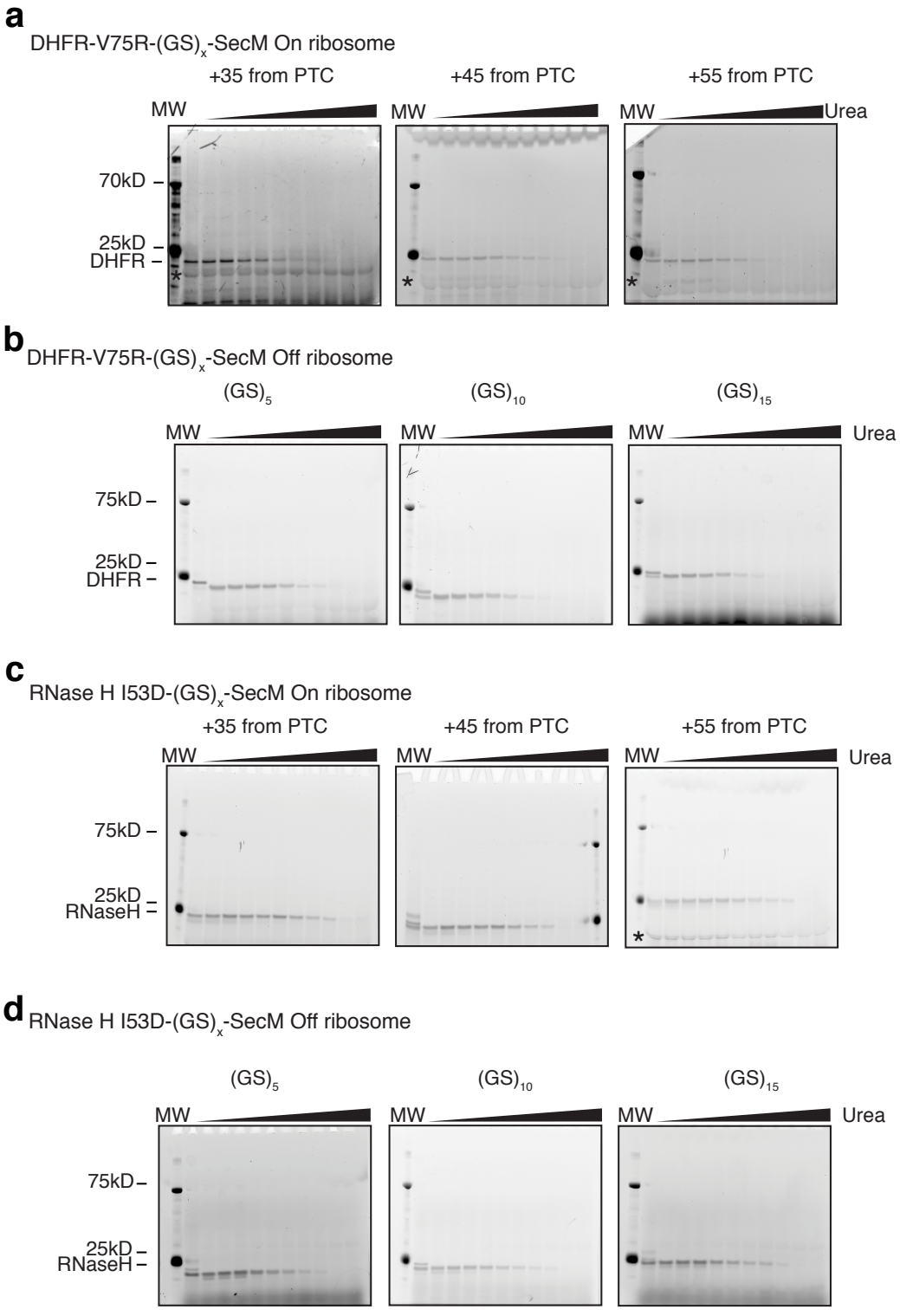


Figure 2.12 - Gels used in Figure 6. (a) DHFR V75R-(GS)x-SecM on ribosome gels. **(b)** DHFR V75R-(GS)x-SecM off ribosome gels. **(c)** RNase H I53D-(GS)x-SecM on ribosome gels. **(d)** RNase H I53D-(GS)x-SecM off ribosome gels.

Table 2.1 – Comparison of IVT and recombinant protein

Protein	$C_{m,purified}$ [Urea] (M)	$\Delta G_{unf,\ purified}$ (kcal/mol)	$C_{m,\ IVT}$ [Urea] (M)	$\Delta G_{unf,\ IVT}$ (kcal/mol)
DHFR V75R	2.50±0.18	4.50±0.32	2.10±0.16	3.78±0.29
RNase H I53D	2.21±0.09	4.41±0.18	2.16±0.10	4.31±0.19

Table 2.2 – Determination of RNC stability

Protein	$C_{m,on}$ [Urea] (M)	$C_{m,off}$ [Urea] (M)	$\Delta G_{unf,on}$ (kcal/mol)	$\Delta G_{unf,\ off}$ (kcal/mol)	$\Delta\Delta G_{unf,\ on\text{-}off}$ (kcal/mol)
DHFR V75R-$(GS)_5$-SecM	1.81±0.05	2.58±0.11	4.57±0.14	6.50±0.25	-1.93±0.29
RNase H I53D-$(GS)_5$-SecM	1.65 ±0.10	2.14±0.13	4.07±0.23	5.29±0.33	-1.22±0.16
Barnase W35F/W94F/H102M-$(GS)_5$-SecM	2.09±0.01	2.31±0.04	3.90±0.02	4.32±0.08	-0.42±0.08

Table 2.3 – Ribosome-mediated destabilization is dependent on distance from the PTC

Protein	Distance from PTC (aa)	$C_{m,on}$ [Urea] (M)	$C_{m,off}$ [Urea] (M)	$\Delta G_{unf,\ on}$ (kcal/mol)	$\Delta G_{unf,\ off}$ (kcal/mol)
DHFR V75R-$(GS)_x$-SecM	35	1.81±0.05	2.58±0.11	4.57±0.14	6.50±0.25
	45	2.47±0.01	2.60±0.15	6.24±0.02	6.55±0.39
	55	2.75±0.02	2.72±0.12	6.93±0.05	6.85±0.30
RNase H-I53D-$(GS)_x$-SecM	35	1.65±0.10	2.14±0.13	4.07±0.23	5.29±0.33
	45	1.92±0.15	2.23±0.11	4.99±0.39	5.80±0.27
	55	2.13±0.11	2.18±0.09	5.96±0.24	5.96±0.30

Chapter 3

Characterization of ribosome nascent chain folding and unfolding kinetics

3.1 Abstract

Protein folding is necessary for proper cellular and organismal function. The disruption of accurate protein folding (misfolding) is highly detrimental to numerous cellular processes and can lead to cell death and disease. A major protein homeostasis (proteostasis) node in the cell is the ribosome, which interacts with a number of quality control factors that help guarantee protein folding fidelity. A protein's initial folding steps can occur during translation, while the protein is still being synthesized. Most studies of a protein's energy landscape, however, are performed in idealized conditions with purified protein in simple buffers. Thus, studies have struggled to replicate conditions that occur during translation and so how the ribosome itself modulates a protein's energy landscape remains unclear. In particular, how the ribosome changes a nascent chain's folding trajectory continues to be a mystery. Here, we measure the unfolding rates of two proteins as stalled nascent chains and find that the unfolding rate of each protein is highly accelerated. This unfolding rate acceleration may serve to add a layer of robustness to the folding process by encouraging the quick resolution of non-native secondary or tertiary structural collapse during translation.

3.2 Introduction

Protein folding is necessary for the function of all cells and protein misfolding is implicated in a wide variety of human diseases. Understanding general physical principles governing protein folding has been a major challenge for physical biochemists and biophysicists over the past half-century. In parallel, biochemists, molecular biologists and cell biologists have worked to uncover how the cell manipulates protein-folding landscapes in order to maintain proteostasis and modulate protein function. The last half-century has been incredibly fruitful for both pursuits; there is a wealth of biophysical data obtained mostly from *in vitro* experiments performed in relatively simple environments with purified proteins[1,2] and scientists have used many *in vivo* and complex *in vitro* techniques to uncover principles governing the maintenance of proteostasis, including the function and regulation of protein chaperones, the translational apparatus, and protein degradation and recycling machineries[3]. Only recently, however, have scientists made major strides unifying these two fields by applying quantitative biophysical techniques to environments more similar to those *in vivo*[4–6]. A major area of research in the past several years has been the study of how the ribosome changes the biophysical properties of the nascent chain: its folding and unfolding rates, stability, and trajectory[7–15]. The surveillance and quality control of nascent chains lies at a proteostasis node in the cell[16,17] and understanding the biophysical characteristics of nascent chain folding will reveal not only general principles of protein folding (*in vivo* and *in vitro*), but also strategies for resolving protein misfolds that can cause cellular insult and are implicated in degenerative diseases including aging, ALS and Alzheimer's[18,19].

With a few notable exceptions[8,11], the study of nascent-chain folding still relies on highly engineered methods that lack throughput and the ability to measure general biophysical properties necessary for comparison to biophysical data generated from both experiment and simulation. In particular, principles for understanding the folding trajectory of the nascent chain and how it is related to its refolding *in vitro* are particularly poorly understood. While many studies have shown that there are differences between folding *in vitro* and during translation, there are still many open questions. Several studies implicate the ribosome's charged surface in modulation protein folding[7,8,13,20], but is this due to binding or repulsion of the nascent chain? Many groups have reported the formation of translation or ribosome-obligate folding intermediates; but do these intermediates form due to the non-equilibrium nature of translation (as suggested by simulation[21–23]) or due to some other property of the translational machinery?

A classic method to probe protein-folding trajectories is to measure the folding and unfolding rates of proteins and their variants in different environments. Unfortunately, measuring protein folding and unfolding during translation is a technical challenge. The presence of the ribosome precludes the use of classic protein spectroscopic techniques, such as fluorescence or circular dichroism (CD), as the signal from the ribosome drowns out that of the nascent chain. Furthermore, translation is a highly asynchronous process– unless folding is measured one molecule at a time[7]–it is

difficult to de-convolute what molecule is folding when during translation. Clearly, we need new biophysical methods that overcome challenges of throughput, detection, time resolution and sample generation.

Here, we combine two extant technologies: (1) the use of the SecM stall sequence to make homogenous stalled Ribosome Nascent Chains (RNCs)[24] and (2) commercially available BODIPY-lysine-peptidyl-tRNA with a previously reported methodology, pulse proteolysis, to measure the unfolding rates of RNCs[8]. We are able to measure the unfolding rate of a two state RNase H variant as well as that of an unfolding intermediate of a DHFR variant. For the variant of RNase H, we find that, compared to off the ribosome, its unfolding rate is vastly accelerated, but the folding rate remains relatively unchanged.

3.3 Pulse proteolysis can be used to measure protein unfolding rates

The biophysical properties of RNCs cannot be measured using standard biophysical techniques and so new types of methods must be used to overcome the challenges associated with studying them (see[2,8,11]). In Chapter 2, we reported the use of pulse proteolysis, combined with an *in vitro* translation and labelling scheme to measure the stability (ΔG_{unf}) of RNCs. Pulse proteolysis is a method to measure the fraction of folded protein in a given condition by taking advantage of the fact that unfolded proteins are more susceptible to proteolysis than folded proteins[25,26]. There are two requirements for accurate measurement of fraction folded using pulse proteolysis: (1) The unfolding rate of the protein must be slower than the pulse length, and (2) The native state must not be cleaved at a similar rate to the unfolded state or access any cleavable excited states. To measure protein stability, pulse proteolysis is performed after equilibration of the target protein at increasing urea concentrations. After the pulse, samples are run on a gel and the band corresponding to the full-length protein is quantified. This band is a direct measurement of the amount of folded protein in solution **(Figure 3.1)**.

Pulse proteolysis can be adapted to measure protein unfolding rates with a simple change in experimental set-up[27]. Briefly, after equilibration in native conditions, unfolding of RNCs or free protein is initiated by dilution into a specific urea concentration. Next, samples are aliquoted into separate tubes and pulse proteolysis is applied at specific time points, in order to measure the fraction of folded protein at that specific time. Samples are then run on a gel and the band corresponding to the full-length protein is quantified and plotted as a function of time. These data are then fit to an exponential:

$$\text{Eq 3.1: } I(t) = I_0 \cdot e^{-kt}$$

Where I_0 is the initial intensity where no protease has been added to the reaction and k is the unfolding rate. It is important to note that observation of a double or higher number of exponential processes would require each unfolding intermediate to have a similar proteolysis rate as the folded state (i.e. pulse proteolysis can only distinguish

between states that are protease susceptible or not within the pulse time). After determining unfolding rates at several different urea concentrations, one can extrapolate to the unfolding rate at zero urea:

$$\text{Eq 3.2: } f(x) = mx \cdot RT^{-1} + k_{unf,H2O}$$

Where RT is the gas constant times temperature and m is the unfolding m-value **(See Figure 3.1)**.

3.4 Using pulse proteolysis to determine the unfolding rate and m-value from *in vitro* translated proteins

Pulse proteolysis has been shown to be a robust method for measuring unfolding rates and unfolding m-values[27] and has successfully recapitulated CD data for the unfolding of RNase H and MBP. To determine if it could be used to measure the unfolding rates of IVT protein, we probed the unfolding of a two-state RNase H variant: I53D/D10A and compared it to that of RNase H I53D/D10A unfolding as measured by CD **(Figure 3.2A and B; Table 3.1)**. We chose RNase H I53D/D10A for two reasons: (1) it is a two-state folder, so by measuring its unfolding rate we can extrapolate its folding rate if we know its stability and (2) it has served as the background for phi-value analysis in previous studies from the Marqusee Lab[28], which we want to repeat on the ribosome. Fitting of the unfolding data revealed a rate constant of $1.6\pm0.7 \times 10^{-6}$ sec^{-1} and and m-value of 0.43 ± 0.7 $kcal \cdot mol^{-1} \cdot M^{-1}$. Although the rate constant is different than previously reported[28], 8.1×10^{-8}, the m-values are quite similar (0.56 vs 0.43±0.7), most likely due to the different conditions: 25mM HEPES pH 7.5, 150mM KCl, 15mM MgOAc, 0.1mM TCEP (here; HKMT) versus 20mM Na Acetate pH 5.5, 50mM KCl[28] for the CD studies. Thus, we measured unfolding of recombinantly produced and purified RNase H I53D/D10A using pulse proteolysis in HKMT at two different urea concentrations **(Figure 3.2C)**. These points fall directly on the fit obtained using IVT protein. We therefore conclude that pulse proteolysis of IVT accurately measures the unfolding rates of RNase H.

3.5 Stability of RNase H I53D/D10A on and off the ribosome

In order to conduct studies on RNase H I53D/D10A on and off the ribosome, its urea-denaturation midpoint, or C_m, must be lower than the urea concentration needed to break apart SecM-stalled 70S ribosomes, about 3.5M urea (Section 2.5 and Chapter 3 reference 8). Using pulse proteolysis, we examined the stability of both free and SecM stalled RNase H I53D/D10A-$(GS)_5$-SecM. Unfortunately, the C_m of RNase H I53D/D10A on the ribosome was determined to be 3.38±0.10M urea, enough that we could measure its stability, but not low enough to perform unfolding studies, as we would like to go at least 1-1.5M urea above the C_m to maximize both rate and signal change for unfolding studies **(Figure 3.3A; Table 3.2).** RNase H I53D/D10A-$(GS)_5$-SecM's stability change due to the ribosome ($\Delta\Delta G_{on\text{-}off}$) is -1.03 ± 0.03 $kcal \cdot mol^{-1}$, similar, but lower than

that of RNase H I53D-$(GS)_5$-SecM **(Figure 3.3B; Table 3.2)**. Interestingly, RNase H I53D's isoelectric point is less than that (more negative) of I53D/D10A and so is its $\Delta\Delta G_{on\text{-}off}$. This is consistent with the anti-correlation of isoelectric point and $\Delta\Delta G_{on\text{-}off}$ that we observe in Chapter 2 **(Figure 3.4)**. Since RNase H I53D/D10A has a C_m too high to perform unfolding studies on the ribosome (which dissociates into 30S and 50S subunits at about 3.5M urea, see chapter 2), we then performed kinetic studies on the more destabilized RNase H variant I53D.

3.6 Unfolding of RNase H I53D on and off the ribosome

We measured the unfolding rates of RNase H I53D-$(GS)_5$-SecM on- and off- the ribosome using pulse proteolysis. First we compared the unfolding parameters we measured for RNase H I53D off the ribosome to those measured previously in the lab[29] **(Figure 3.5A and B; Table 3.3)**. Similar to RNase H I53D/D10A, the extrapolated unfolding rate was quite different, $6.3\pm5\text{x}10^{-6}$ sec^{-1} (Spudich et al.) versus $2.95\pm1.9\text{x}10^{-5}$ sec^{-1} (here), but the m-values were very similar (this is identical to the case of RNase H I53D/D10A). This is not unexpected, as RNase H I53D in conditions here is destabilized by 1.4 ± 0.10 $kcal\cdot mol^{-1}$ compared to in the previous study. Since RNase H I53D is two-state, this destabilization must arise from a change in either the folding or unfolding rates. Here, it is primarily due to changes in the unfolding rates as the equilibrium constant is reduced 10-fold and the unfolding rate is reduced 10-fold. It is important to note that this stability change is not due to the C-terminal addition of the glycine-serine linker or SecM sequence (see Figure 2.1). Next, we measured the unfolding rate of RNase H I53D as a stalled RNC **(Figure 3.5C and D and table 3.3)**. We found the unfolding rate to be accelerated by one log order, accounting for almost all of the stability difference measured between the protein as a RNC and free in solution. The m-values measured on the ribosome are within error of those determined off the ribosome. This contrasts with a previous study which suggests that the folding rate and not unfolding rate is primarily effected by the ribosome[7].

3.7 Unfolding of DHFR V75R on and off the ribosome

In order to generalize our results further, we applied pulse proteolysis to the unfolding of DHFR V75R both on and off the ribosome. DHFR V75R-(GS)5-SecM is destabilized by 1.93 ± 0.29 kcal/mol due to its stalling as a nascent chain[8]. DHFR, however, is known to fold and unfold through several intermediates which could preclude analysis of DHFR's unfolding rate by pulse proteolysis. Consistent with this, unfolding of DHFR V75R-(GS)5-SecM's (off the ribosome) yields different results when measured by pulse proteolysis versus CD. When measured by pulse proteolysis, unfolding is slower by a factor of ~5.7 **(Figure 3.6A-D and Table 3.4)**. Interestingly, the unfolding m-values also differed, by a factor of two. The m-value as measured by CD is half that as measured by pulse proteolysis, suggesting that the CD experiments are reporting on unfolding to an intermediate not accessible by proteolysis which reports on yet another unfolding intermediate. These two phases, each about one log order

between them, have been observed multiple times for WT DHFR. Most DHFR folding studies are done using intrinsic fluorescence and some states of DHFR and its mutants are differentially visible to CD spectroscopy[30–33]. We hypothesize that DHFR V75R-(GS)5-SecM undergoes a fast unfolding phase that thermolysin cannot cleave, but is visible by CD. We then proceeded to measure DHFR V75R-(GS)5-SecM's unfolding rate on the ribosome and found it to be only slightly different than that measured off the ribosome ($k_{unf,off}$ = 2.1±4.2x10^{-4}±0.7; $k_{unf,on}$ = 5±1.0x10^{-4}±0.2). The m-values both on and off the ribosome are, just as in the RNase H case, indistinguishable from each other **(Table 3.4)**. This does not mean, however, that there is no acceleration of the unfolding rate, just that pulse proteolysis cannot measure the slower phase, as the transition to the cleavable species occurs faster.

3.8 Discussion

Here, we have demonstrated that pulse proteolysis can be used for determining unfolding rates of stalled RNCs. For RNase H I53D, we observe a drastic acceleration in the unfolding rate that accounts for almost all of the change in stability due to stalling on the ribosome. For DHFR V75R, we capture a slow unfolding step that differs from a faster unfolding step that is visible to CD and the difference in rate on and off the ribosome is minimal. This does not mean that there is no unfolding rate acceleration due to the ribosome. We cannot rule out, however, that the change in DHFR V75R RNC stability is due to deceleration of the protein's folding rate. In fact, this hypothesis makes sense structurally, as the *in vitro* refolding pathway of DHFR requires docking of its C-terminal beta-sheet to complete folding. Since this sheet may be sterically hindered by the ribosome exit tunnel at short linker distances, it is possible that this is the rate-limiting step and hence the folding rate, and not the unfolding rate, is primarily affected.

Equally of interest to the extrapolated unfolding rates measured here is the unfolding m-value. There are not enough points to accurately obtain an equilibrium m-value with pulse proteolysis, but because there are (1) fewer parameters when fitting unfolding and (2) each point on the unfolding limb is made up of many data points, unfolding m-values can be accurately calculated using pulse proteolysis[34]. Here, we observe no change in m-values whether on or off the ribosome and the m-values we measure for RNase H I53D and RNase H I53D/D10A are within error of those measured using CD. This is in agreement for comparison of force dependent m-values on- and off- the ribosome for T4 lysozyme[7]. Since m-values are related to the change in exposed surface area upon unfolding, our results suggest that the destabilization we report in Chapter 2 and the differences in rate measured here for RNase H I53D are not due to the nascent chain binding to the ribosome. If the destabilization of the nascent chain were due to favoring an unfolded, but ribosome-bound state, then we would expect a decrease in the m-value of the nascent chain, as the change in accessible surface area upon folding would decrease. Another possibility is that the ribosome bound state can only be accessed from the unfolded state, meaning that folding/unfolding would proceed as follows:

$$U_{ribosome\text{-}bound} \leftrightarrow U_{free} \leftrightarrow F$$

If the ribosome-bound and free unfolded states were both fully susceptible to proteolysis during the one-minute pulse, or if the exchange between $U_{ribosome\text{-}bound}$ and U_{free} were very fast, then it is possible we would not observe a change in m-value.

What could be the advantage to accelerating unfolding during translation versus decelerating folding? During translation, the nascent chain has time to sample conformations it otherwise would not, as the full sequence of the protein has not yet been synthesized. The energetics of hydrophobic collapse, however, make burial of hydrophobic side chains quite favorable. The nascent chain could, then, potentially collapse to kinetically trapped translational intermediates. Interestingly, only on-pathway co-translational intermediates or the abrogation of off-pathway intermediates have been observed on stalled nascent chains or co-translationally[35–37]. The cell therefore has two choices: decrease the folding rate so potential hydrophobic collapse occurs more slowly, with the hope that enough sequence emergences so that the protein can avoid toxic states; or to accelerate the resolution of those states in the first place. Since the destabilization we have measured so far is on the order of 10-fold, and translation is at least 100-fold slower than secondary and tertiary structure formation, decelerating folding would not provide enough time for peptide synthesis to occur to avoid collapse. This then leaves us with accelerating the unfolding of potential misfolds. Data in Chapter 2 and this chapter suggest that this is done by decreasing the energy difference between the folded and unfolded states (thus lowering the barrier to the unfolded state). Still, the nature of the unfolding rate acceleration we observe here and the destabilization reported in Chapter 2 remains unknown. With these techniques in hand, however, we are now poised to begin answering deeper questions about co-translational folding, the ribosome, and eventually, protein folding in the cell.

3.9 Materials and Methods

Sample preparation:
50μL IVT reactions without release factors (PURExpress, NEB) were initiated by addition of 500ng plasmid DNA in the presence of 2.5μL Flourotect Greenlys tRNA (Promega) and 2μL RNase inhibitor, murine (NEB).

Preparation of samples off the ribosome or without a SecM sequence:
Samples were incubated for 1 hr at 37°C before adding chloramphenicol to 2mM and RNase A to a final concentration of 0.1 mg/mL. These samples were then incubated at room temperature overnight and spun at 21000xg for 30 minutes at 4°C. The supernatant was then used for pulse proteolysis.

Preparation of RNCs:
After incubation for 30 minutes to 1 hour at 37°C, IVT reactions were loaded onto a 125μL 1M sucrose cushion in 25mM HEPES pH 7.5, 15mM MgOAc, 150mM KCl, 2mM DTT (HKM+DTT) and centrifuged at 200,000xg for 40 minutes at 4°C. Supernatant was

aspirated and ribosome pellets were washed three times with 200μL HKM+DTT. Pellets were then resuspended in 35μL of HKM +DTT (see Figure 2.9).

Pulse Proteolysis:
For protein purified from *E. coli*, pulse proteolysis was conducted as previously described in HKM+DTT. For released or stalled nascent chains, 22.5μL of sample was mixed into a final volume of 170μL in HKM+DTT to the desired urea concentration to initiate refolding. Samples were then aliquoted into 10μL increments and pulsed with 1μL of 6.8mg/mL Thermolysin for one minute. 8μL of reaction was then quenched into a new tube with 3μL of .5M EDTA. pH 8.0. After pulse proteolysis, RNase A was added to 0.1mg/mL to each reaction and incubated at 37°C overnight to digest any remaining peptidyl-tRNA. For IVT reactions off the ribosome, RNase A was added to a final concentration of 0.1mg/mL and incubated 15 minutes at 37°C. Samples were then mixed with SDS-PAGE loading dye and loaded onto 4-12% Bis-Tris gels (ThermoFisher, Waltham, MA). Gels were run in MES buffer and imaged with a typhoon (GE, Schenectady, NY) using a 488nm laser and 520BP filter. Analysis and quantification of gels was done using ImageJ as described previously[25,26]. Urea concentrations were measured using a refractometer as described previously[25,26]. Fitting was done in Matlab.

CD

For unfolding experiments, protein was incubated in HKM+0.1mM TCEP (HKMT) overnight at room temperature. Protein was diluted 1:20 into HKMT+urea to reach the desired final urea concentration. All experiments were performed with manual mixing at a protein concentration of 100μg/mL in a .5cm cuvette and monitored at a wavelength of 225nm to increase signal to noise. Analysis of data was performed as in[29]. Fitting was done in Matlab.

3.10 References

1. Braselmann E, Chaney JL, Clark PL. Folding the proteome. *Trends Biochem Sci*. 2013;38(7):337-344. doi:10.1016/j.tibs.2013.05.001.
2. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. *Curr Opin Struct Biol*. 2010;20(1):33-45. doi:10.1016/j.sbi.2010.01.005.
3. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591.
4. Ebbinghaus S, Dhar A, McDonald JD, Gruebele M. Protein folding stability and dynamics imaged in a living cell. *Nat Methods*. 2010;7(4):319-323. doi:10.1038/nmeth.1435.
5. Smith AE, Zhou LZ, Gorensek AH, Senske M, Pielak GJ. In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci U S A*. 2016;113(7):1725-1730. doi:10.1073/pnas.1518620113.

6. Kim SJ, Yoon JS, Shishido H, et al. Translational tuning optimizes nascent protein folding in cells. *Science*. 2015;348(6233):444-448. doi:10.1126/science.aaa3974.
7. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C. The ribosome modulates nascent protein folding. *Science*. 2011;334(6063):1723-1727. doi:10.1126/science.1209740.
8. Samelson AJ, Jensen MK, Soto RA, Cate JHD, Marqusee S. Quantitative determination of ribosome nascent chain stability. *Proc Natl Acad Sci*. 2016;113(47):201610272. doi:10.1073/pnas.1610272113.
9. Holtkamp W, Kokic G, Jäger M, Mittelstaet J, Komar AA, Rodnina M V. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104-1107. doi:10.1126/science.aad0344.
10. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858-861. doi:10.1021/ja411302m.
11. Nilsson OB, Nickson AA, Hollins JJ, et al. Cotranslational folding of spectrin domains via partially structured states. *Nat Struct & Mol Biol*. 2017;24(3):221-226. doi:10.1038/nsmb.3355.
12. Nilsson OB, Hedman R, Marino J, et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep*. 2015;12(10):1533-1540. doi:10.1016/j.celrep.2015.07.065.
13. Cabrita LD, Cassaignau AME, Launay HMM, et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol*. 2016;(February). doi:10.1038/nsmb.3182.
14. Eichmann C, Preissler S, Riek R, Deuerling E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2010;107(20):9111-9116. doi:10.1073/pnas.0914300107.
15. Hingorani KS, Gierasch LM. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Curr Opin Struct Biol*. 2014;24:81-90. doi:10.1016/j.sbi.2013.11.007.
16. Pechmann S, Willmund F, Frydman J. The ribosome as a hub for protein quality control. *Mol Cell*. 2013;49(3):411-421. doi:10.1016/j.molcel.2013.01.020.
17. Tyedmers J, Mogk A, Bukau B. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol*. 2010;11(11):777-788. doi:10.1038/nrm2993.
18. Hartl FU. Protein Misfolding Diseases. *Annu Rev Biochem*. 2017;(SEPTEMBER):228-234. doi:10.1146/annurev-biochem-061516-044518.
19. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426(6968):884-890. doi:10.1038/nature02261.
20. Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, Cavagnero S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem Biol*. 2013;8(6):1195-1204. doi:10.1021/cb400030n.
21. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc*. 2011;133(3):513-526. doi:10.1021/ja107863z.
22. O'Brien EP, Vendruscolo M, Dobson CM. Prediction of variable translation rate

effects on cotranslational protein folding. *Nat Commun*. 2012;3:868. doi:10.1038/ncomms1850.
23. Ciryam P, Morimoto RI, Vendruscolo M, Dobson CM, O'Brien EP. PNAS Plus: In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proc Natl Acad Sci*. December 2012. doi:10.1073/pnas.1213624110.
24. Evans MS, Ugrinov KG, Frese MA, Clark PL. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods*. 2004;2(10):757–762. doi:10.1038/NMETH790.
25. Park C, Marqusee S. Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr Protoc Protein Sci*. 2006;Chapter 20:Unit 20.11. doi:10.1002/0471140864.ps2011s46.
26. Park C, Marqusee S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*. 2005;2(3):207-212. doi:10.1038/nmeth740.
27. Na Y-R, Park C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci*. 2009;18(2):268-276. doi:10.1002/pro.29.
28. Connell KB, Miller EJ, Marqusee S. The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms. *J Mol Biol*. 2009;391(2):450-460. doi:10.1016/j.jmb.2009.05.085.
29. Spudich GM, Miller EJ, Marqusee S. Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol*. 2004;335(2):609-618. doi:10.1016/j.jmb.2003.10.052.
30. Jennings PA, Finn BE, Jones BE, Matthews CR. A Reexamination of the Folding Mechanism of Dihydrofolate Reductase from. 1993:3783-3789.
31. Touchette N, Perry K, Matthews C. Folding of dihydrofolate reductase from Escherichia coli. *Biochemistry*. 1986:5445-5452. http://pubs.acs.org/doi/abs/10.1021/bi00367a015. Accessed October 8, 2014.
32. Tronrud DE, Eyck LF Ten, Matthews BW. An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures. 1986;(November).
33. Garvey EP, Matthews CR. Effects of multiple replacements at a single position on the folding and stability of dihydrofolate reductase from Escherichia coli. *Biochemistry*. 1989;28(5):2083-2093. http://www.ncbi.nlm.nih.gov/pubmed/2655702.
34. Na Y, Park C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci*. 2009;18:268-276. doi:10.1002/pro.29.
35. Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*. 1999;6(7):697-705. doi:10.1038/10754.
36. Houwman JA, André E, Westphal AH, van Berkel WJH, van Mierlo CPM. The Ribosome Restrains Molten Globule Formation in Stalled Nascent Flavodoxin. *J Biol Chem*. 2016:jbc.M116.756205. doi:10.1074/jbc.M116.756205.
37. Evans MS, Sander IM, Clark PL. Cotranslational folding promotes beta-helix

formation and avoids aggregation in vivo. *J Mol Biol*. 2008;383(3):683-692. doi:10.1016/j.jmb.2008.07.035.

3.11 Figures

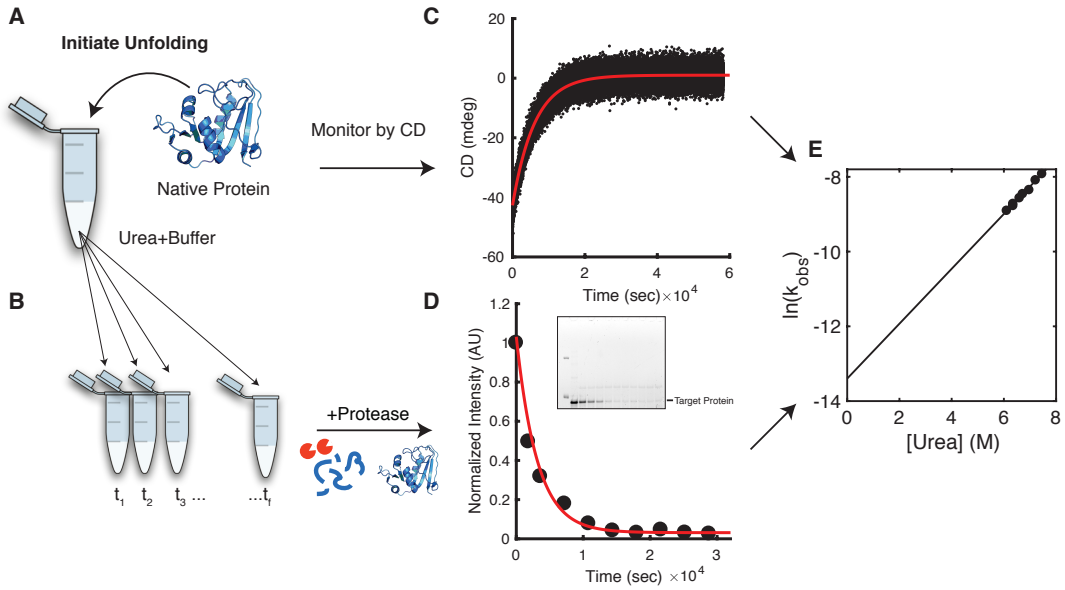


Figure 3.1 – Pulse proteolysis can be used to measure unfolding rates. (a) Protein unfolding is initiated by diluting native protein into a urea containing solution**. (c)** Traditionally, this unfolding reaction would be observed through some spectroscopic method over time and a rate would be calculated and plotted as a function of urea concentration. For pulse proteolysis **(b**), the unfolding reaction is aliquoted and protease is added at specific time points and subsequently quenched **(d)**. After the reaction is complete, each time-point is run on a gel and a rate is calculated. This is repeated at various urea concentrations and the log of the rate is plotted as a function of urea concentration to determine the unfolding rate in no urea and the unfolding m-value.

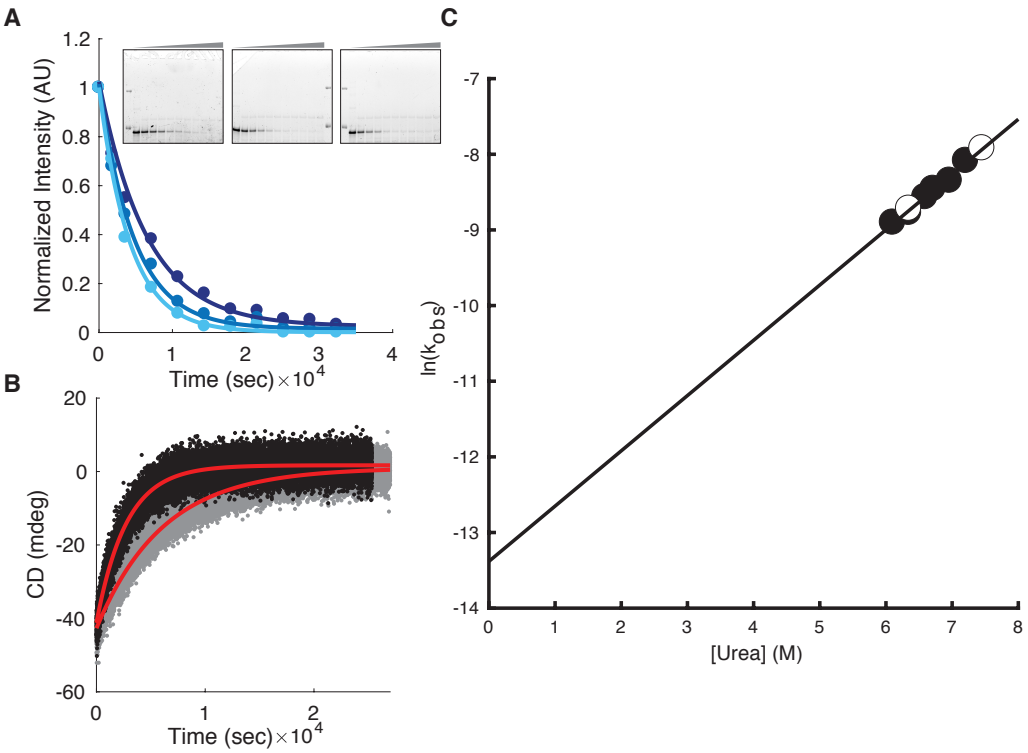


Figure 3.2 – Pulse proteolysis of *in vitro* translation protein accurately recapitulates unfolding rates of purified protein using CD. (A) Unfolding traces at different urea concentrations as obtained by pulse proteolysis. **(B)** Unfolding traces obtained by CD at two different urea concentrations. **(C)** Plot of the natural log of the unfolding rate as a function of urea. Fit and black circles are obtained using pulse proteolysis of *in vitro* translated protein, white circles with black outline are CD unfolding data with purified protein and are not used to generate the fit.

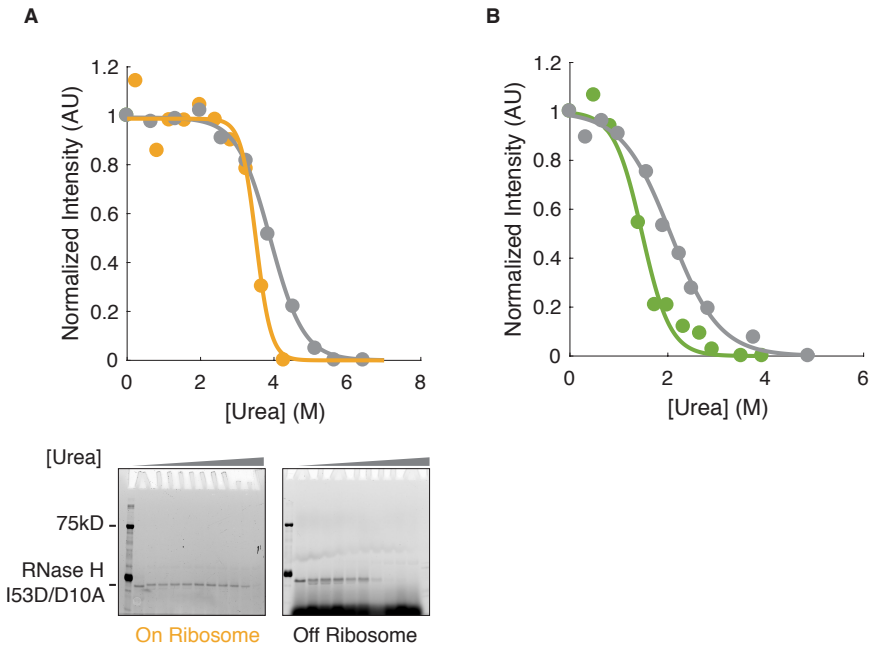


Figure 3.3 – The stabilities of RNase H I53D and RNase H I53D/D10A are decreased due to ribosomal tethering. (A) RNase H I53D/D10A both on (orange) and off (grey) the ribosome. Lines are fits to the data. **(B)** RNase H I53D both on (green) and off (grey) the ribosome. Lines are fits to the data. **(C)** Gels of RNase H I53D/D10A data on and off the ribosome. Gels for I53D are available in Figure 2.5. Gels and traces are representative of three separate experiments.

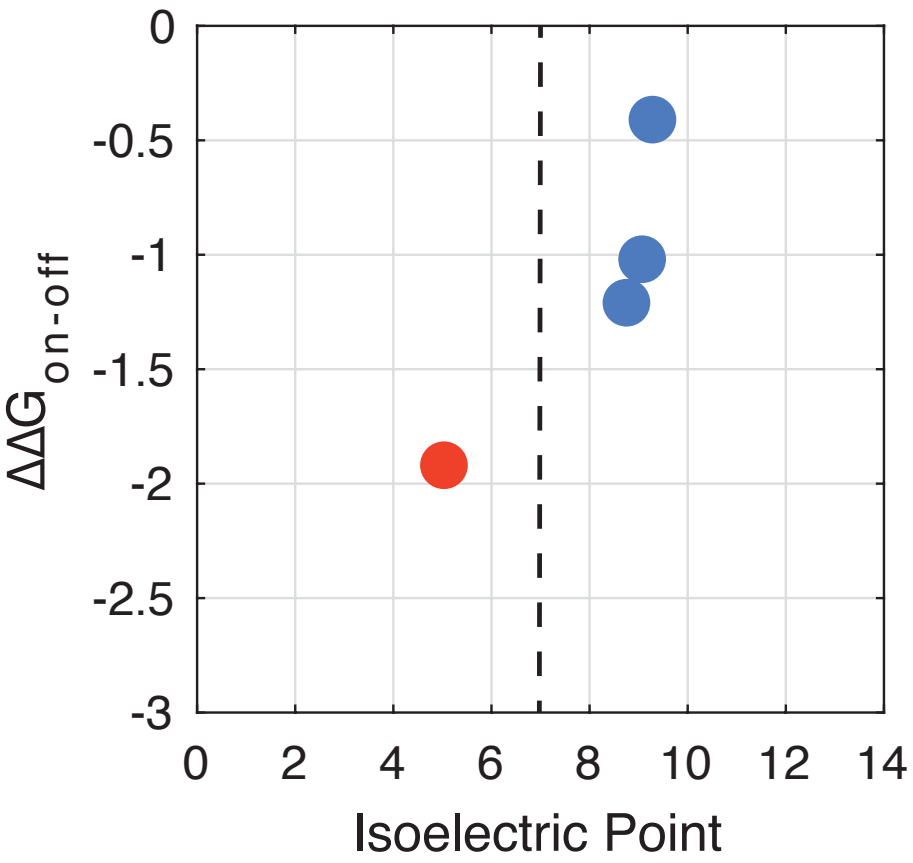


Figure 3.4: Plot of protein isoelectric point versus change in stability due to the ribosome. Proteins with pIs under 7.0 are in red, proteins with pIs above 7.0 are in blue. Dashed line is at pI = 7.0.

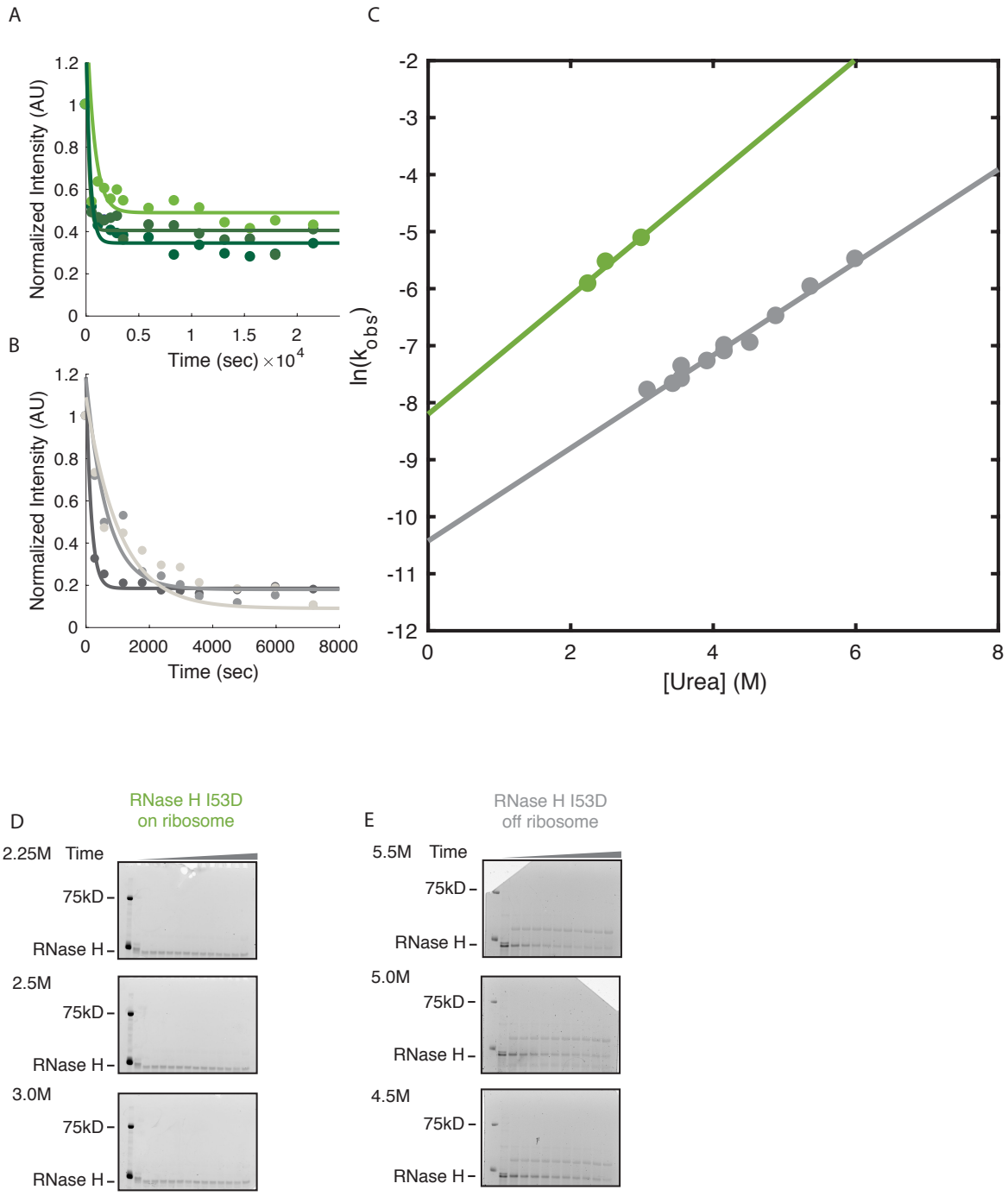


Figure 3.5 – Unfolding of RNase H I53D both on and off the ribosome as determined by pulse proteolysis. (A) Unfolding experiments at a final concentration of 2.0 (dark green), 1.75 (green), 1.5 (light green) molar urea. **(B)** Unfolding experiments at a final concentration of 5.5 (dark grey), 5.0 (grey), 4.5 (light grey) molar urea**. (C)** Plot of the natural log of the observed rate versus urea concentration. Green (on ribosome) and grey (off ribosome). **(D) and (E)** Gels for data shown in **(A**) and **(B)**.

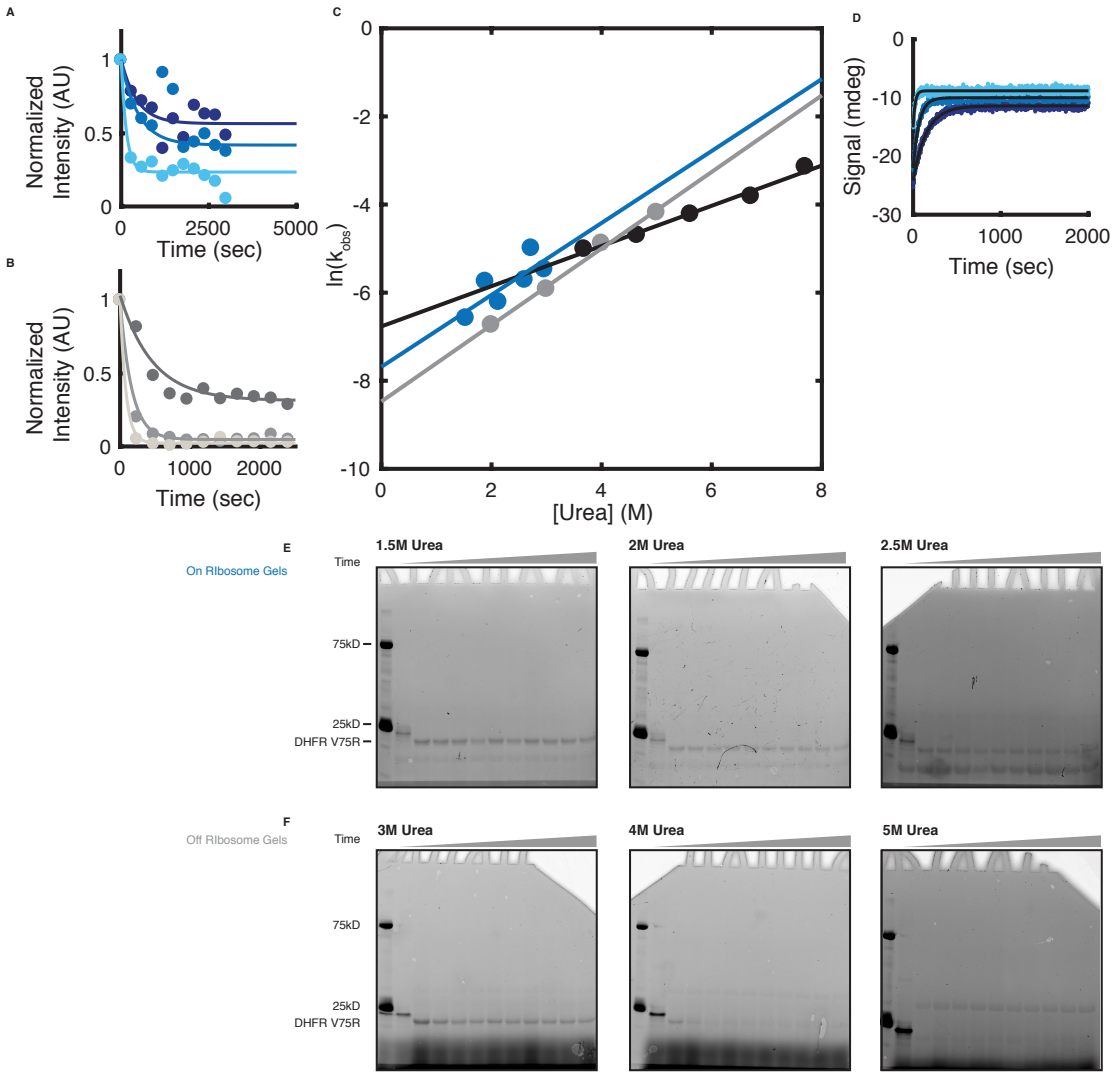


Figure 3.6 – Unfolding of DHFR V75R both on and off the ribosome determined by pulse proteolysis and CD. (A) Unfolding experiments at a final concentration of 1.5 (dark blue), 2 (blue), 2.5 (light blue) molar urea. **(B)** Unfolding experiments at a final concentration of 5 (dark grey), 4 (grey), 3 (light grey) molar urea**. (C)** Plot of the natural log of the observed rate versus urea concentration. Green (on ribosome) and grey (off ribosome). **(D)** Unfolding experiments at a final concentration of 3.5 (dark blue), 4.5 (blue), 5.5 (light blue) molar urea **as measured by CD.** Three **(E) and (F)** Gels for data shown in **(A**) and **(B)**.

3.12 Tables

Table 3.1 – Unfolding parameters of RNase H I53D/D10A determined by CD (Connell et al.[28]) and pulse proteolysis

	k_{unf} (sec^{-1})	m_{NU} ($kcal \cdot mol^{-1} \cdot M^{-1}$)
CD (Connell et al.[28])	8.1×10^{-8}	0.56
Pulse Proteolysis	$1.6 \pm 0.70 \times 10^{-6}$	0.43±0.7

Table 3.2 – Determination of stability for RNase H I53D-$(GS)_5$-SecM and RNase H I53D/D10A-$(GS)_5$-SecM both on and off the ribosome

	$C_{m,on}$ (urea, M)	$C_{m,off}$ (urea, M)	$\Delta\Delta G_{on-off}$ (kcal·mol^{-1})
RNase H I53D-$(GS)_5$-SecM	1.65±0.10	2.14±0.13	-1.22±0.16
RNase H I53D/D10A-$(GS)_5$-SecM	3.38±0.10	3.87±0.04	-1.03±0.03

Table 3.3 – Determination of the unfolding rates of RNase H I53D-$(GS)_5$-SecM both on and off the ribosome

	k_{unf} (sec^{-1})	m_{NU} ($kcalmol^{-1}M^{-1}$)
$I53D_{cd}$ (Spudich et al.)[29]	$6.3\pm5x10^{-6}$	0.5±0.1
$I53D_{off}$	$2.95\pm1.9x10^{-5}$	0.48±.06
$I53D_{on}$	$2.74\pm9.1x10^{-4}$	0.61±1.5

Table 3.4 – Determination of the unfolding rates of DHFR V75R-$(GS)_5$-SecMboth on and off the ribosome

	m ($kcal \cdot mol^{-1} \cdot M^{-1}$)	k_{unf} (sec^{-1})
CD	0.27±0.07	$1.1 \pm 0.14 \times 10^{-3}$
Pulse $Proteolysis_{off}$	0.51±0.12	$2.1 \pm 0.1 \times 10^{-4}$
Pulse $Proteolysis_{on}$	0.48±0.40	$5.0 \pm 4.1 \times 10^{-4}$

Chapter 4

The ribosome vastly re-orders HaloTag folding to avoid an aggregation-prone folding intermediate and increase folding efficiency

4.1 Abstract

Precise protein folding is necessary for the function of most proteins, and inaccurate protein folding (misfolding) is implicated in a wide-range of diseases. For many proteins, folding begins during synthesis, and it has recently become appreciated that protein quality control at the ribosome is essential for maintaining cellular proteostasis. With only a handful of exceptions, however, the current biophysical view of nascent chain folding is based either on experiments in the absence of translation (stalled nascent chains) or low-resolution studies that do not describe general biophysical properties of proteins. Thus, there exists both a lack of detailed mechanistic data for how steps in the folding process change during translation and a lack of biophysical data that are directly comparable to state of the art protein-folding simulations and experiments. Here, we develop and utilize a new, high-throughput model system for monitoring protein folding kinetics during translation using the protein HaloTag. Furthermore, we structurally and kinetically characterize HaloTag's folding pathway during both translation and refolding. We find that translation abrogates a fast-forming, aggregation-prone intermediate that is otherwise present during refolding and demonstrate that translation itself is enough to change HaloTag's folding pathway and increase its folding efficiency. These results further our understanding of how the interplay between folding and translation can guarantee the folding fidelity of the proteome.

4.2 Introduction

Protein folding is essential for the function of almost every protein in the cell and protein misfolding can cause cell death and disease[1]. Ever since Anfinsen demonstrated that the information for a protein's fold is contained in its primary sequence, the role of translation and its implications for protein folding have been under investigation[2–5]. However, detailed mechanistic information about the folding process is usually obtained from *in vitro* refolding studies: unfolding proteins using denaturant and then observing their refolding *while the entire polypeptide sequence is available*. This has the potential to be very different from co-translational folding [6–9] because the timescale for secondary and tertiary structure formation (msec-sec) is much shorter than translation (minutes), the developing nascent chain has the opportunity to explore unique regions of the energy landscape. Furthermore, many long-range native contacts (such as beta-sheets) are prohibited from forming in the early stages of translation as the amino acids involved are either not yet synthesized or are still within the ~100Å long ribosome exit tunnel.

Indeed, in the last two decades, co-translational folding has been implicated in changing nascent chain folding efficiency, folding speed, and quaternary structure formation when compared to in vitro refolding from denaturant or protein assembly from individually purified subunits[10–15]. While it is clear that direct interactions with specific chaperones play an important role in the *in-vivo* folding process, understanding the process of translation and the ribosome itself also has the potential to reveal general and ubiquitous strategies for protein quality control. For instance, intermediates that form uniquely during co-translational folding are implicated in the efficient folding of several proteins (e.g. Fedorov and Baldwin, 1999; Frydman et al., 1999) and ribosome-obligate intermediates have been detected[17,18]. In light of recent advances towards understanding a protein's trajectory towards toxic aggregates[19–22], this presents a conundrum: translation allows the nascent chain time to populate partially-folded intermediates; at the same time, such partially folded states are often themselves on-pathway toward toxic aggregates. Indeed, inhibition of nascent chain quality control machinery in eukaryotic cells can lead to the formation of large inclusions[23–25].

How do proteins avoid non-native traps during translation? Does ribosomal occlusion of residues separated in sequence space, but not in structure, prevent aggregation that would otherwise occur during *in vitro* refolding experiments or after ribosomal release? To address these questions and investigate the role of the ribosome and the process of translation in protein folding, we need new tools that go beyond probing the conformations of stalled nascent chains, and monitor the kinetics and conformations of the nascent chain folding trajectory during translation.

Here, we develop and utilize new tools to monitor cotranslational folding of the protein Halotag in order to probe the role of co-translational folding and its potential to alter a protein's folding pathway to help avoid aggregation-prone intermediates. HaloTag is a modified haloalkane dehalogenase from *Rhodococcus sp*. *(* Promega (Madison, WI)) and is commonly used as a tool for *in vivo* imaging. HaloTag presents a unique and highly-multiplexable model system; it is a 35 kilodalton (kD) single domain

protein engineered to covalently bind ligands functionalized with a variety of flourophores and other small molecules. We first demonstrate that HaloTag is a viable model system for measuring co-translational folding rates using fluorescence polarization (FP). Next, we use both circular dichroism and pulse-labeling hydrogen exchange to characterize the refolding trajectory of HaloTag. Then, by monitoring both processes (co-translational folding and traditional refolding) using pulse proteolysis, we determine that *de novo* folding of HaloTag is more efficient than refolding. Finally, we demonstrate structural differences between these two folding processes by monitoring cysteine accessibility during both refolding and *in vitro* translation. We conclude that the different conformations sampled during these two folding processes alter its folding efficiency.

4.3 HaloTag refolding can be monitored by fluorescence polarization

In its native conformation, HaloTag covalently binds commercially available ligands functionalized with a variety of fluorescent dyes. The reported second order rate constant for ligand binding is ~ 2.7×10^{-6} $M^{-1}sec^{-1}$ [26], corresponding to a rate of ~2.7 sec^{-1} at 1μM-5μM protein (the protein concentrations used here). Thus, if the folding of HaloTag is much slower than 2.7 sec^{-1}, we can monitor the concentration of folded protein by measuring the amount of functionalized ligand (hereafter called TMR-ligand) bound to the protein. To measure the amount of TMR-ligand bound to HaloTag as a function of time, we use fluorescence polarization (FP), which reports on the relative tumbling time (overall dimensions) of the measured flourophore. **Figure 4.1C** shows the dramatic increase in the fluorescence polarization signal when TMR-ligand is mixed with folded (0M Urea) HaloTag as compared to unfolded HaloTag (8M Urea), consistent with selective covalent modification of native protein. It is important to note we do not observe any transient associated with the signal change when we add native protein, confirming that the binding reaction is complete within the dead-time of our experiment (10 sec). Next, we rapidly diluted the unfolded HaloTag (8M Urea) 10-fold into folding conditions (0.8 M Urea) in the presences of TMR-ligand to see if HaloTag refolding is slower than the rate of ligand binding (~2.7 sec^{-1}). Under these conditions, HaloTag folds at a rate of $4.79 \pm 0.61 \times 10^{-4}$ sec^{-1} ($k_{f,H2O,FP}$) or about 10,000 times more slowly than ligand binding **(Figure 4.1D, Table 4.1)**, as measured by FP. This was confirmed by circular dichroism (see below). Thus, the amount of TMR-ligand-Halo complex is directly proportional to the amount of folded protein at any time during the refolding reaction and folding can be monitored using fluorescence polarization.

4.4 Co-translational folding can be monitored by fluorescence polarization

The wavelength of TMR-ligand fluorescence (576nm) is far away from the intrinsic fluorescence of most biomolecules. Therefore, unlike most other standard measurements of protein folding, such as CD or tryptophan fluorescence, the folding of HaloTag can be monitored in the presence of other proteins and biomolecules. Thus, we applied the same fluorescence polarization strategy to monitor HaloTag folding

during *in vitro* translation. In this case, instead of diluting from high urea, a coupled *in vitro* transcription/translation system was initiated by the addition of plasmid DNA containing the HaloTag sequence under control of a T7 promoter directly in the fluorimeter. The kinetic trace **(Figure 4.2A)** shows two phases: (1) a lag phase where there is no increase in polarization and (2) an exponential phase, where the fluorescence polarization increases as a function of time and reaches a maximum. Fitting these data yields a rate of $4x10^{-4}$ sec^{-1}, very close to the rate observed when folding is monitored after dilution from urea. Herein we will refer to refolding from dilution simply as refolding and co-translational folding as folding or *de novo* folding). As a control, the same experiment was carried out using the DNA encoding the protein DHFR, which should not bind the TMR ligand; *in vitro* translation of DHFR shows a very slow linear increase in the polarization signal, likely due to evaporation over time. Importantly, for HaloTag, only the final polarization value and not the rate is dependent on the concentration of TMR-ligand **(Figure 4.2B, Table 4.2)**.

4.5 The co-translational folding rate is not limited by translation

The above co-translational experiments monitor the production of native HaloTag. Is the observed rate related to protein folding or protein production? To address this, we monitored the time dependence of protein production using a gel assay: we quenched *in vitro* translation reactions performed in the presence of tRNA-BODIPY-Lysine (Promega; Mallam and Jackson, 2012) with RNase, EDTA, and chloramphenicol at specific time points and determined the amount of full-length HaloTag by in-gel fluorescence **(Figure 4.2A and D)**. Translation displays a similar two-phase behavior to the change in polarization signal, but the translation lag phase is significantly shorter than the folding lag phase. This lag phase represents the time it takes to synthesize a detectable amount of unfolded protein and is 275±57.2 seconds long, corresponding to an average translation rate of about one amino acid per second, slower than *in vivo* translation, but similar to translation rates in other *in vitro* systems [28–30]. In comparison, the lag time observed for the change in polarization is an order of magnitude larger than the translation lag time, 2009.3±72.7 seconds **(See Table 4.2)**. In addition, we also halted translation by the addition of Neomycin, an antibiotic that inhibits translation, at a time were there should be significant amounts of unfolded protein. If increases in FP were due to folding, and not translation, then we would expect to observe an increase in FP even after the addition of Neomycin. This is precisely what we observe after addition of neomycin after 4500 seconds of translation time **(Figure 4.2C)**. Thus, we can conclude that folding is not limited by the rate of translation and the change in fluorescence polarization is due to folding of HaloTag. Interestingly, the exponential phases of both translation and folding are similar in rate, although clearly translation occurs much earlier than folding.

4.6 HaloTag aggregates when refolded < 1.0M Urea

We next characterized the folding landscape of HaloTag using circular dichroism in

order to compare refolding and *de novo* folding. Equilibrium urea-induced denaturation revealed a single cooperative unfolding process. Fitting this transition using a two state, linear extrapolation yielded a ΔG_{unf} = 6.03±0.39 kcal mol^{-1} and m-value of 1.57±0.1kcal mol^{-1} M^{-1}**(Figure 4.3D and 4.3E)**. It is interesting to note that this m-value is about two-fold lower than that expected for a protein of 35kD[31]. Often a smaller than expected m-value is a sign that the process is not two state, with the presence of a hidden equilibrium intermediate, which can often be detected by monitoring with a different probe. However, monitoring equilibrium denaturation with intrinsic fluorescence (following tertiary structure as compared to secondary structure) revealed the same equilibrium m-value and stability as those measured using CD **(Figure 4.4)**.

We next examined the kinetics of refolding and unfolding using CD. **Figure 4.3** shows the resulting chevron plot (the observed rates as a function of [Urea]). Unfolding experiments revealed an apparent rollover at high concentrations of urea, evidence that an intermediate may be populated during unfolding **(Figure 4.3C,G)**. There is also evidence of a burst-phase in the unfolding regime **(Figure 4.3G)**, but global fitting to a three-state model did not significantly reduce uncertainty compared to a two-state kinetic model **(See Table 4.1)**. Folding studies revealed two phases, a fast-folding phase and a urea independent slow phase **(Figures 4.3A-C, F-G)**. Often, urea-independent folding phases are due to cis-trans proline isomerization. However, refolding and *in vitro* translation folding experiments in the presence of the proline isomerase Cyclophilin A revealed no change in the measured folding rate **(Figure 4.5)**. To our surprise, HaloTag did not refold efficiently below 1.0M urea, displaying altered folding kinetics and visible precipitation after at least 300 seconds of refolding time **(Figure 4.3A)**. It is important to note that at urea concentrations of under 1.0M, aggregation occurred after an initial decrease in signal intensity that has a similar rate to the fast refolding phase observed in non-aggregating conditions **(Figure 4.3A)**.

4.7 Co-translational folding is more efficient than refolding

Since refolding to a final concentration below 1.0M Urea resulted in visible aggregation **(Figure 4.6),** we wondered if such aggregation occurred during cotranslational folding and decided to determine the folding efficiency of the two processes. First, to determine the amount of protein that forms aggregate we separated the soluble and insoluble fractions of the refolding reaction using centrifugation, resuspended the insoluble pellet in the original volume of 8M urea and determined the fraction of native protein in the supernatant versus that in the pellet using in-gel fluorescence. Using this approach, we find that 70±6% of the protein remains in the soluble fraction **(Figure 4.7A)**. To determine the fraction of folded protein after refolding, we turned to pulse proteolysis[32].

Pulse proteolysis is a gel-based method for measuring the amount of folded protein in solution that takes advantage of the fact that unfolded proteins are more susceptible to proteolysis than folded proteins. Protease (here, thermolysin) is added to an equilibrium mixture of protein for an amount of time sufficient to degrade all the unfolded proteins, but short enough that folded proteins remain intact. Then, the

sample is run on a SDS-PAGE gel and the band corresponding to the full-length protein is a direct measurement of the amount of folded protein in a given condition as long as the protein unfolding-rate is slower than the pulse length.

When we performed pulse proteolysis on HaloTag brought from zero to 0.8M urea, we found no change in the amount of folded protein. In contrast, when Halo was refolded from 8.0M to 0.8M urea, the fraction of folded protein as measured by pulse proteolysis matched that of the amount in the supernatant after centrifugation– 0.73±0.3 versus 0.70±.0.06 respectively **(Figure 4.7B)**.

Is cotranslational folding also inefficient? To evaluate the folding efficiency of HaloTag after *in vitro* translation, we initiated IVT reactions, quenched them with chloramphenicol and RNase A after two hours and after equilibration overnight assessed the amount of folded HaloTag in solution using pulse proteolysis. In contrast to HaloTag refolding, we found *de novo* folding to be significantly more efficient ($p<.01$, student's unpaired t-test; n= 15): 0.91±0.03 versus 0.70±.06 **(Figure 4.7B, Table 4.3)**. This effect is unlikely due to differences in protein concentration between refolding and *in vitro* translation reactions as the protein concentration during *in vitro* translation is greater than the protein concentration during refolding (>5μM and 3μM respectively; see **Figure 4.2B**). To rule out differences between *in vitro* translated protein and protein produced in *E. coli*, we refolded *in vitro* translated protein and determined its refolding efficiency to be more similar to that of purified protein (0.69±0.06; **Figure 4.7B**). Additionally, refolding by dialysis also resulted in visible aggregates and a fraction of folded protein equal to 0.74±0.1 (**Figure 4.7B)**. Since there are no chaperones present during *in vitro* translation and there is still an observable transient before aggregation takes place as observed by CD, we hypothesized that the aggregation could be due to self-association of a fast-forming intermediate and that the ribosome could be occluding distant regions of the protein which form interactions necessary for formation of this intermediate. We then proceeded to characterize which regions of HaloTag fold first using hydrogen exchange mass spectrometry (HX/MS).

4.8 Following the refolding trajectory using HX-MS

To monitor the refolding trajectory of HaloTag at a structural level, we followed refolding using HX-MS under conditions where no aggregation is detected (a final urea concentration equal to 1.6M). Refolding of HaloTag was initiated by dilution of deuterated protein into 1.6M urea in HKMT at 10C. At various refolding times, a short (10ms) pulse of high pH is used to initiate hydrogen exchange followed by a quench to pH 2 at 10C. After quenching, the protein was immediately fragmented with pepsin and fungal protease and subjected to LC/MS. If an amide-deuterium is unprotected during the pulse, it will exchange with hydrogen resulting in an apparent mass decrease of peptides containing that residue. Thus, changes in the mass of a specific peptide is a measure of backbone amide accessibility at that particular time-point in refolding[33,34]. Although we obtained 360 peptides with greater than 98% coverage fo the protein, we only used peptides which were present at all time points in our analysis, limiting our dataset to 104 peptides. These peptides, however, still have >90% coverage of the

entire protein. **Figure 4.8A** shows the median fraction deuterated for each peptide at each time point. The peptides can be separated into two populations, those which are at least 50% deuterated by 10 seconds after initiation of refolding ("fast", red) and those that are not ("slow", blue). This distinction is maintained at later time points and a larger proportion of "fast" peptides have plateaued by 300 seconds than "slow" peptides. Plotting the protection of each peptide as a function of time also clearly reveals these two separate populations **(Figure 4.8B).** We then mapped these peptides onto the structure of HaloTag.

The structure of HaloTag, like many dehalogenases, is made up of, a Rossman fold, which serves as a structural scaffold for its active site, with a large, completely alpha-helical insertion between residues 129 and 236 that forms a lid over the active site which confers substrate specificity. Almost the entire Rossman fold is protected within ten seconds of refolding (ß-strands 1-3, 5-7 and α-helices A, C, I and L) **(Figure 4.8A-C)**. In contrast, the entire lid domain (residues 130-235, α-helices D-I) as well as ß-strand 8 are protected much more slowly **(Figure 4.8A-E)**. These data suggest that the fast-folding phase we observe by CD is due to collapse of most of the Rossman fold and that the slow folding phase is the folding and docking of HaloTag's lid domain.

4.9 The folding pathway of HaloTag is altered during translation

The HXMS experiments outlined above are not feasible for studies of cotranslational folding due to the large pH jumps required for labeling. Therefore, to compare the refolding trajectory determined above with the cotranslational folding trajectory of HaloTag we measured cysteine accessibility during both IVT and refolding experiments. We monitored cysteine accessbility using a fluorescein conjugated malemiede (FSM) that can be read out using in-gel fluorescence. This method has two advantages: (1) HaloTag contains two buried cysteines that are completely protected in the folded state **(Figure 4.1 and Figure 4.9A and B)** and (2) Fluorescence can be read out on an SDS-PAGE gel, providing both high sensitivity and separation of all labeled products by size.

HaloTag's two cysteines are positioned at the base of two beta-strands, ß4 and ß8. In refolding monitored by HXMS, ß4 shows intermediate behavior – it is protected more slowly that the fast folding phase, but faster that the slow folding phase; ß8 is part of the slow-folding phase. Both cysteines are good probes of folding – they are completely protected in the folded state and fully labelled within 30 seconds in the unfolded state **(See Figure 4.9 and Figure 4.10)**. We probed these cysteines during refolding in non-aggregation conditions using a time course of 30-second pulses with FSM **(Figure 4.9A).** The data were best fit using a double exponential, with a fast phase of .03±.02 sec^{-1} and a slow phase of 7.8±0.6x$10^{-4}$$sec^{-1}$ **(See Table 4.1)**. These values are within two-fold of HaloTag's fast folding (0.04±0.02 sec^{-1}) and slow folding rates (6.6±.71x10^{-4} sec^{-1}) as determined by CD. These same cysteine residues were monitored during co-translational folding: we initiated an IVT reaction, halted further translation with the addition of chloramphenicol after 45 minutes, and probed cysteine accessibility with FSM as a function of time **(Figure 4.9A)**. Unlike the time course

observed during refolding, cysteine protection followed a single exponential indistinguishable from HaloTag's slow folding rate during refolding **(See Table 4.1)**. There was no evidence for a fast-folding stage involving these cysteines.

We then created site-specific cysteine residues to probe the very early stages of folding: Halo** E121C, Halo** I126C and Halo** M129C. Halo** is a cysteine-free version of Halo where its two native cysteines have been mutated to alanine. All three variants bind TMR and display similar folding kinetics as WT HaloTag **(See Figure 4.11)**. Residues 126 and 129 are both buried and located on ß6, which HX-MS reveals to be protected before 10 seconds of folding time **(See Figure 4.8A-C)**. Halo** E121C probes the surface of HaloTag and serves as a positive control. Again, we followed the time trajectory of protection for all three sites during both refolding and co-translational folding. During refolding, both Halo** I126C and M129C are protected within the burst phase of the experiment **(**black dots, **Figure 4.9B-C)**. The final level of protection (20% of unfolded) is consistent with that expected from studies on the folded state **(Figure 4.12)**. (I126C and M129C are only protected to 20% of unfolded in 1.6M urea). Thus, cysteine accessibility is a good probe for the collapse of ß6 as it recapitulates the HXMS data. In contrast, Halo** E121C remains unprotected throughout the entire time trajectory **(Figure 4.9D)**. Thus, the time course of protection of Halo** I126C and M129C reports on the collapse of an early intermediate during HaloTag refolding. We next probed Halo** E121C, I126C and M129C cysteine protection during *in vitro* translation. Surprisingly, cysteines in Halo** I126C and M129C do not show immediate protection and instead show a time-dependent transient with a similar rate to that of overall HaloTag folding (as in WT, compare **Figures 4.9 B-C to 6A, and see Table 4.1**). Again, as expected, Halo** E121C, is continually accessible **(Figure 4.9D)**. From these data, we conclude that translation prevents the formation of an early, collapsed folding intermediate seen during *in-vitro* refolding and is likely responsible for the increased co-translational folding efficiency (see below).

4.10 Discussion

Here, we have shown that the ribosome and the vectorial process of translation can modulate the folding efficiency of HaloTag by changing the folding trajectory of the protein. When HaloTag is refolded, it populates a fast forming intermediate that involves residues distant in sequence space contacting each other. Furthermore, a significant portion of the protein remains unfolded in the possible presence of a large exposed hydrophobic surface. Helix B and ß-sheet 4 are unprotected during formation of the Rossman fold core despite the fact they both make critical contacts with the ß1, ß2, and αC **(Figure 4.8E)**. Furthermore, peptides corresponding to these regions do not fall clearly within either "slow" or "fast" categories **(Figure 4.8B)**. We hypothesize this is because ß4 and αB are connected by a >15 residue linker that forms contacts with the slow-folding αI, which must dock in order for there to be full protection. Thus, during the early steps of folding not only is there a large segment of the protein that is completely unprotected (αD-I, the lid domain), but there is also a large amount of exposed hydrophobic surface area (ß1-3 and αC). We hypothesize that the formation of most of

the Rossman fold before the docking of either the lid domain or ß4 and αB provides an exposed hydrophobic surface that leads to aggregation during refolding. During translation, however, the Rossman fold cannot form, as it involves residues on both the N-terminus and C-terminus of the lid domain, meaning that the entire lid must emerge from the ribosome and (by extension) have time to fold before the Rossman fold can completely collapse. Thus, docking of the lid and ß4 and αB4 could occur concurrently or before formation of the Rossman fold, therefore inhibiting aggregation **(Figure 4.13)**. Unfortunately, characterization of transient folding intermediates during *in vitro* translation remains a challenge and we lack structural insight into HaloTag's on-ribosome folding intermediates.

The idea that translation may change the trajectory or even the final conformation of a protein is not new and our data is consistent with several other studies exploring how translation changes the folding efficiency and/or rate of protein folding [10,15,35,36]. It is interesting to consider our data in the light of several new studies which show that the ability to access transient intermediates is a major determinant for the formation of toxic aggregates associated with protein variants that cause neurodegenerative disease [19,21,22,37]. Our data suggest that the process of translation may add a layer of robustness to the folding process by preventing the formation of aggregation-prone conformations.

In addition to having a unique folding landscape, HaloTag is also a promising model system for studying co-translational folding. Since we now know which secondary structural elements fold when, it should be possible to compare HaloTag's folding trajectories not just to those during translation but in the presence of chaperones and other biomolecules. Furthermore, we have reported here that HaloTag folding can be measured by ligand binding and thus folding experiments can be performed with high-throughput and in the presence of many other biologically active molecules. This will enable groups to systematically investigate the translational machinery and determine each component's effect on HaloTag folding. Hopefully these types of unbiased approaches will lead to the discovery of general and quantitative rules that govern not only protein folding during translation, but protein folding in other high complexity environments.

4.11 Materials and Methods

HaloTag and mutant growth and purification

Growth

BL21(DE3) cells were transformed with expression vectors containing the wild-type or mutant HaloTag cDNA. Single colonies were used to seed starter cultures grown overnight to saturation. Large scale cultures were inoculated with 5mL of overnight culture, grown at 37 C to an OD600 of 0.6-0.8 and induced with 1 mM IPTG for 2-3 hours at 37 C. After induction, cultures were pelleted at 5000 x g for 10 minutes at 4C, flash frozen and stored at -80C.

Purification
Cell pellets were resuspended in 10 mM Tris/H_2SO_4, pH 7.5, 1 mM TCEP (Lysis Buffer) and lysed by sonication on ice. Lysates were cleared by centrifugation for 30 minutes at 20,000 x g, 4°C and subsequently filtered through 0.2 μm filters. After clearing, lysate was dialyzed into at least a 10-fold volume excess of Lysis Buffer, loaded onto a HiPrep 16/10 Q Xl column equilibrated with Lysis Buffer and eluted with a gradient of Lysis bBuffer plus 0 to 600 mM NaCl. Fractions containing the HaloTag protein were dialyzed into into at least a 10-fold volume excess of 20 mM Sodium Acetate, pH 5.0 (Q Buffer) loaded onto a HiPrep 16/10 Q Xl column equilibrated with Q Buffer and eluted with a gradient of Q Buffer plus 0 to 800 mM NaCl. Fractions containing HaloTag protein were then concentrated and purified on a HiLoad 16/600 Superdex 75 pg column equilibrated with 50 mM ammonium bicarbonate or 25mM HEPES KOH pH 7.5, 15mM MgOAc, 150mM KCl, 0.1mM TCEP (HKMT) and the fractions with the retention volume corresponding to the size of monomeric HaloTag were either lyophilized (ammonium bicarbonate runs) and subsequently stored at -80C or concentrated and immediately used for experiments (HKMT runs). All lyophilized protein was resuspended in HKMT and spun filtered before use in experiments.

Fluorescence Polarization

Data Collection
All experiments were performed at 37C unless otherwise noted. Fluorescence polarization was performed on a BioTek Ssynergy Nneo2 plate reader in 384-well, black flat bottom plates for IVT reactions (Corning) or 96 well clear flat bottom plates (refolding experiments). Acquisitions were collected using polarizers as well as a 530nm/590nm filters with side gain set at 45 and top gain set at 40. Read height was 7.5mm and 10 measurements were made per data point. After a 15 minute incubation at 37C, readings initialized by the addition of DNA or ribosomes (IVT reactions) or unfolded protein (refolding experiments) were taken every 20 seconds for five hours after thirty seconds of mixing and a 90 second delay for temperature equilibration.

IVT Reactions
IVT reactions were set up on ice as per the manufacturer's protocols for a 30μL reaction with the addition of 1μL RNase inhibitior, Murine and 1uL of 300μM TMR (in 100% anhydrous DMSO, for a final concentration of 10μM) and pipetted into wells. Plates were covered with clear titer-tops to prevent evaporation and equilibrated at 37°C for 15 minutes. Reaction were initiated with 2μL of 125ng/μL plasmid DNA.

Refolding Experiments
Refolding experiments were performed in HKMT buffer plus appropriate concentrations of urea and TMR added to 5μM so the concentration of DMSO was 3.33%. Plates were sealed and incubated at 37C for fifteen minutes until reactions were initiated by adding 10μL of 20μM HaloTag in 8M urea that had been incubated at 37°C for at least 12 hours. Refolding traces were fit to the following equation in Matlab, using bi-square

fitting and "*k*" bounded at zero:

$$f(t) = a*(1-e^{-kt})+c$$

Urea concentrations were measured using a refractometer as previously described (ref).

Circular Dichroism

Kinetic and equilibrium were performed using a 0.5cm cuvette at 37C with constant stirring at 3μM (0.1mg/mL) in HKMT buffer. Equilibrium and kinetic experiments were performed as previously described[38], but at a wavelength of 225nm instead of 222nm to increase signal-to-noise. Analysis was performed as described[38].

Wavelength experiments were performed in a 0.1cm cuvette at 37°C with 15μM protein (0.5mg/mL) in HKMT buffer).

Determination of Folding Efficiency

All reactions were performed at 37°C at a final concentration of 3μM protein in HKMT buffer unless otherwise noted.

Centrifugation assay
Proteins were refolded by the dilution of protein in 8M urea to the proper urea concentration and allowed to reach equilibrium for at least 12 hours. Samples were then centrifuged at 21130xg for 30 minutes and the supernatant was carefully removed. The pellet was resuspended in an equal volume of 8M urea. Both the supernatant and pellet were then mixed with 6x SDS-PAGE loading dye and run on a 4-12% Bis Tris gel in MES run buffer and subsequently stained with SYPRO Red for 30 minutes in 10% acetic acid. After destain in 10% acetic acid or an hour, gels were imaged using a Typhoon Trio (GE) and analyzed with imageJ.

Pulse Proteolysis
IVT reactions were performed as per the manufacturers instructions but with the addition of 1uL of RNase inhibitor, murine and 1.25uL Flourotect Greenlys (Promega) per 25uL IVT reaction. IVT reactions were quenched after 1 hour to a final concentration of 2mM Chloramphenicol and 0.1mg/mL RNase A. Refolding experiments were performed as described above. IVT reactions and refolding reactions were allowed to reach equilibrium for at least 12 hours. Subsequently, reactions were aliquoted to 10uL and 1uL of 1mg/mL Thermolysin (Sigma) was added to each reaction for 1 minute and quenched with EDTA to a final concentration of 83mM. SDS-PAGE loading dye was then added to each reaction and then run on a 4-12% Bis Tris gel in MES run buffer. Imaging and analysis was performed as described previously[39].

Refolding of IVT translated protein

IVT reactions were performed and quenched as described above. A 10-fold volume excess of 8M urea in HKMT buffer was then added and mixed to the IVT translation reaction and allowed to equilibrate at 37C overnight. Reactions were then concentrated in a 0.5mL 10kD cutoff spin concentrator (Amicon) and diluted to 0.8M urea. After equilibration at 37C overnight, pulse proteolysis was performed as described.

Translation Rate Measurement

IVT reactions were performed as per the manufacturer's instructions but with the addition of 1μL of RNase inhibitor, murine and 1.25μL Flourotect Greenlys (Promega) per 25μL IVT reaction and initiated with 250ng/25μL IVT reaction of DNA. At each time point, 1.5μL of IVT reaction was quenched into a final concentration of 2mM Chloramphenicol and 0.1mg/mL RNase A and then SDS-PAGE loading dye. Reactions were then run on a 4-12% Bis Tris gel in MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ.

Cysteine protection assays

Purified protein
Refolding reactions were initiated as described above. At each time point, a 50-fold excess of flourescein-maleimide (FSM; supplier) was added for 30 seconds and quenched into an equal volume of SDS-PAGE loading dye containing Beta-mercaptoethanol to a final BME concentration of 2.15M. Reactions were then run on 4-12% Bis Tris gel in MES run buffer and imaged using a Typhoon Trio. Analysis was performed using ImageJ. Traces were fit to the following equation in Matlab, using bi-square fitting and "k" bounded at zero (for those data which displayed exponential kinetics):

$$a*(1-e^{-kt})+c$$

Since Halo**E121C cysteine reactivity is time-dependent over the labelling time of the reaction, intensities after refolding was initiated were normalized to the reactivity at that labelling time.

IVT reactions
IVT reactions were initiated as described above. At each timepoint, an equal volume of 2mM FSM was mixed with IVT reaction for 30 seconds and quenched into SDS-PAGE loading dye as above. At 45 minutes, reactions were halted by the addition of chloramphenicol to a final concentration of 2mM. Reactions were then run on a 4-12% Bis Tris gel in MES run buffer and imaged using a Typhoon Trio. A sample of labeled HaloTag was run to determine the size of the Halo-Flourescein band marked with a '*' in Figure 4.9. Analysis was performed using ImageJ. Intensities were normalized to a major protein product running at ~65kD to control for effects of evaporation, fluorescein bleaching and gel loading. Traces were fit to the following equation in Matlab after exclusion of points before 45 minutes, using bi-square fitting and "*k*" bounded at zero:

$$f(t) = a*(1-e^{-kt})+c$$

Pulsed Labeling HX-MS

QFM refolding reaction protein

The HX-MS pulsed labeling scheme and mass spec measurement used here is similar to previously described approaches (1). Deuterated protein was prepared by 4 cycles of drying protein in 8M Urea and resuspension with D_2O. For refolding and pulsed labeling reactions, a Bio-logic QFM-4 is used in an interrupted flow mode to refold and quickly label unfolded regions of the protein. To initiate refolding, deuterated protein in high-denaturant buffer (8M urea in HKMT) is diluted with 10 volumes of deuterated HKMT to a final urea concentration of 1.6M. After a variable delay time to allow refolding, a high-pH pulse of proteated buffer (200 mM Glycine, 10ms, 5 volumes; final pH=10.00 final solution is 31% proteated) used to label unstructured regions of the protein with ^1H. The pulse is then quenched by dilution with a low pH buffer (1M Glycine, 5 volumes; final pH=2.00) to slow the exchange reaction. Protein samples are then collected and injected into a custom LC/MS system. A folded control sample was prepared by subjecting deuterated, native protein to the same pulse/quench sequence, and an unfolded control was measured by performing the pulse/quench on fully deuterated, unfolded protein.

LC/MS system

A custom HPLC system was used for in-line protease digestion, desalting, and separation of peptides. Peptides were eluted from the trap column and separated on an analytical C8 column using an acetonitrile gradient at 17 μL/min. The output of this system was directly injected into a Thermo Scientific LTQ Orbitrap Discovery using electrospray ionization.

Data Analysis of HX-MS pulsed labeling

Peptides were identified using a SEQUEST search using Proteome Discoverer 2.0 software. Peptide mass envelopes were fit using HDExaminer (Sierra Analytics) followed by manual confirmation of each peptide. Deuterium content was assessed by examination of the centroid of each fitted peptide mass envelope. Only peptides with high signal/noise ratio at each timepoint were used for further analysis. The raw number of deuterons protected in each peptide were normalized to folded and unfolded control samples. Each peptide protection curve was fit to either a single exponential or sum of exponentials using Matlab and rates were compared between secondary structures using the crystal structure of HaloTag (PDB 5UY1).

4.12 Acknowledgements

We would like to thank the entire Marqusee lab for advice on experiments, paper writing and support especially S.A. Lim for help with kinetic fitting and MKJ for help with the paper. Also, we would like to thank A. Martin and his lab for equipment and reagents

and especially D. Costanzo for advice relating to FSM labelling. J.S. Fraser for advice on the paper. J.Chen for technical support.

4.13 References

1. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426(6968):884-890. doi:10.1038/nature02261.
2. Evans MS, Clarke TF, Clark PL. Conformations of co-translational folding intermediates. *Protein Pept Lett*. 2005;12(2):189-195. http://www.ncbi.nlm.nih.gov/pubmed/15723645.
3. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591.
4. Clark PL. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci*. 2004;29(10):527-534. doi:10.1016/j.tibs.2004.08.008.
5. Braselmann E, Chaney JL, Clark PL. Folding the proteome. *Trends Biochem Sci*. 2013;38(7):337-344. doi:10.1016/j.tibs.2013.05.001.
6. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc*. 2011;133(3):513-526. doi:10.1021/ja107863z.
7. Rodnina M V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci*. 2016;25:1390-1406. doi:10.1002/pro.2950.
8. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. *Curr Opin Struct Biol*. 2010;20(1):33-45. doi:10.1016/j.sbi.2010.01.005.
9. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol*. 2009;16(6):589-597. doi:10.1038/nsmb.1614.
10. Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*. 1999;6(7):697-705. doi:10.1038/10754.
11. Kolb VA, Kommer A, Spirin AS. Co-Translational Protein Folding in Prokaryotic and Eukaryotic Cell-Free. 2002;7(95).
12. Gloge F, Becker AH, Kramer G, Bukau B. Co-translational mechanisms of protein maturation. *Curr Opin Struct Biol*. 2014;24:24-33. doi:10.1016/j.sbi.2013.11.004.
13. Shieh Y-W, Minguez P, Bork P, et al. Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science*. 2015;350(6261):678-680. doi:10.1126/science.aac8171.
14. Fedorov AN. Cotranslational Protein Folding - Minireview. *J Biol Chem*. 1997;272(52):32715-32718. doi:10.1074/jbc.272.52.32715.
15. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009;16(3):274-280. doi:10.1038/nsmb.1554.
16. Fedorov AN, Baldwin TO. Process of biosynthetic protein folding determines the rapid formation of native structure. *J Mol Biol*. 1999;294(2):579-586. doi:10.1006/jmbi.1999.3281.

17. Holtkamp W, Kokic G, Jäger M, Mittelstaet J, Komar AA, Rodnina M V. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104-1107. doi:10.1126/science.aad0344.
18. Conti BJ, Elferich J, Yang Z, Shinde U, Skach WR. Cotranslational folding inhibits translocation from within the ribosome-Sec61 translocon complex. *Nat Struct Mol Biol*. 2014;21(3):228-235. doi:10.1038/nsmb.2779.
19. Lim KH, Dasari AKR, Hung I, et al. Solid-State NMR Studies Reveal Native-like β-Sheet Structures in Transthyretin Amyloid. *Biochemistry*. 2016;55(37):5272-5278. doi:10.1021/acs.biochem.6b00649.
20. Lim KH, Dasari AKR, Hung I, Gan Z, Kelly JW, Wemmer DE. Structural Changes Associated with Transthyretin Misfolding and Amyloid Formation Revealed by Solution and Solid-State NMR. *Biochemistry*. 2016;55(13):1941-1944. doi:10.1021/acs.biochem.6b00164.
21. Karamanos TK, Pashley CL, Kalverda AP, et al. A Population Shift between Sparsely Populated Folding Intermediates Determines Amyloidogenicity. *J Am Chem Soc*. 2016;138(19):6271-6280. doi:10.1021/jacs.6b02464.
22. Jahn TR, Radford SE. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys*. 2008;469(1):100-117. doi:10.1016/j.abb.2007.05.015.
23. Choe Y-J, Park S-H, Hassemer T, et al. Failure of RQC machinery causes protein aggregation and proteotoxic stress. *Nature*. 2016;531(7593):191-195. doi:10.1038/nature16973.
24. Shen PS, Park J, Qin Y, et al. Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. SUPPLEMENT. *Science*. 2015;347 VN-(6217):75-78. doi:10.1126/science.1259724.
25. Zhou C, Slaughter BD, Unruh JR, et al. Organelle-based aggregation and retention of damaged proteins in asymmetrically dividing cells. *Cell*. 2014;159(3):530-542. doi:10.1016/j.cell.2014.09.026.
26. Los G V, Encell LP, McDougall MG, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol*. 2008;3(6):373-382. doi:10.1021/cb800025k.
27. Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol*. 2012;8(2):147-153. doi:10.1038/nchembio.742.
28. Lodish HF, Jacobsen M. Regulation of hemoglobin synthesis. Equal rates of translation and termination of - and -globin chains. *J Biol Chem*. 1972;247(11):3622-3629. http://www.ncbi.nlm.nih.gov/pubmed/4624122.
29. Guet CC, Bruneaux L, Min TL, et al. Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res*. 2008;36(12). doi:10.1093/nar/gkn329.
30. Zubay G. In vitro synthesis of protein in microbial systems. *Annu Rev Genet*. 1973;7(5):267-287. doi:10.1146/annurev.ge.07.120173.001411.
31. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity

changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci*. 1995;4(10):2138-2148. doi:10.1002/pro.5560041020.
32. Park C, Marqusee S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*. 2005;2(3):207-212. doi:10.1038/nmeth740.
33. Hu W, Walters BT, Kan Z, et al. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A*. 2013;110(19):7684-7689. doi:10.1073/pnas.1305887110.
34. Mayne L, Kan ZY, Sevugan Chetty P, Ricciuti A, Walters BT, Englander SW. Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J Am Soc Mass Spectrom*. 2011;22(11):1898-1905. doi:10.1007/s13361-011-0235-4.
35. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858-861. doi:10.1021/ja411302m.
36. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. December 2012. doi:10.1038/nsmb.2466.
37. Hulleman JD, Balch WE, Kelly JW. Translational attenuation differentially alters the fate of disease-associated fibulin proteins. *FASEB J*. 2012;26(11):4548-4560. doi:10.1096/fj.11-202861.
38. Spudich GM, Miller EJ, Marqusee S. Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol*. 2004;335(2):609-618. doi:10.1016/j.jmb.2003.10.052.
39. Samelson AJ, Jensen MK, Soto RA, Cate JHD, Marqusee S. Quantitative determination of ribosome nascent chain stability. *Proc Natl Acad Sci U S A*. November 2016. doi:10.1073/pnas.1610272113.

4.14 Figures

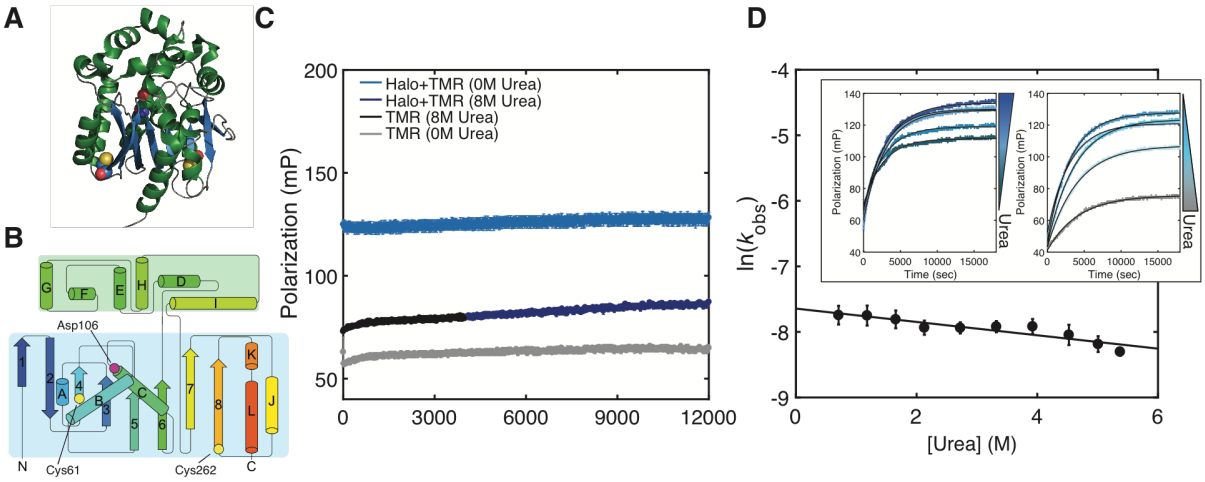


Figure 4.1 – HaloTag folding can be measured using fluorescence polarization. **(A)** X-ray crystal structure of HaloTag (RCSB: 5UY1). Helices colored green, ß-sheets colored blue. In spheres are HaloTag's two cysteines (residues 61 and 262) as well as the residue which forms a covalent bond with HaloTag ligands (aspartic acid 106). **(B)** Secondary structure representation and labelling of HaloTag colored rainbow from N to C-terminus. In blue box, is the core Rossman fold which makes up almost all haloalkane dehalogenases. In green box is HaloTag's lid domain, which is completely alpha-helical. **(C)** TMR fluorescence polarization is increased only in the presence of folded HaloTag (light blue). TMR in 0M urea is in grey; TMR in 8M urea is in black. Unfolded HaloTag in the presence of TMR in 8M urea is in dark blue. **(D)** Folding of HaloTag as measured by fluorescence polarization to increasing urea concentrations (black circle and line). Error bars are ±SD of three experiments. **Inset left** Raw data showing the change in fluorescence polarization over time. Dark Turquoise is 0.7M urea and increasing concentrations of urea are shown until 2.7M urea in dark blue. **Inset Right** Raw data showing the change in fluorescence polarization over time. Light blue is 3.3M urea and traces of increasing urea concentration move down the color gradient until grey, at 5.4M urea.

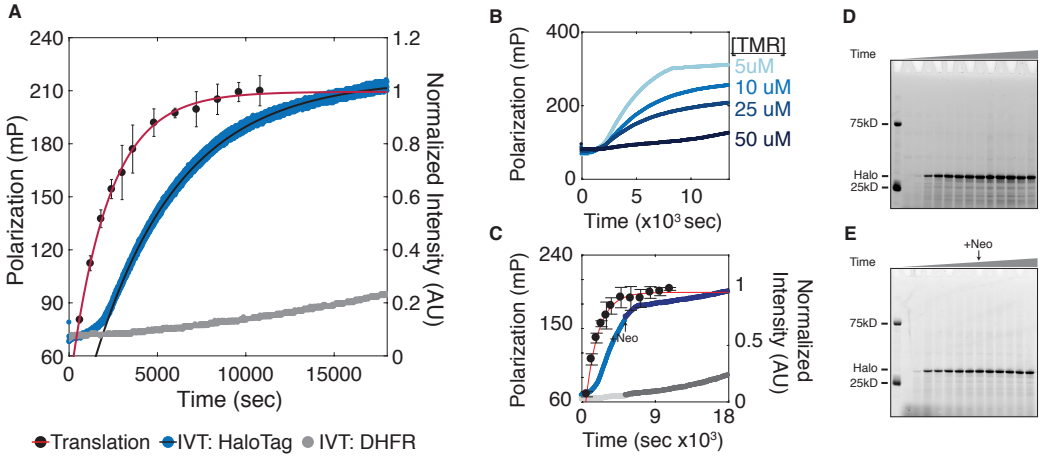


Figure 4.2 - HaloTag folding can be measure during *in vitro* translation. (A) Flourescence polarization of HaloTag during *in vitro* translation (blue circles, black line) increases as a function of time significantly after translation (black circles, with red line). In grey is the change in fluorescence polarization as a function of time during *in vitro* translation of DHFR which does not bind TMR-ligand. **(B)** Flourescence polarization as a function of TMR concentration shows a difference in final intensity but not in rate (also see Table 1). **(C)** Flourescence polarization signal continues to increase even after translation is halted by the addition of Neomycin at 5400 seconds. Data before the addition of neomycin is shown in light blue, data after neomycin addition is shown in dark blue. Translation is shown in black with a red line. Errorbars for each panel represent the SD of three separate experiments. If not shown, error is smaller than each point.

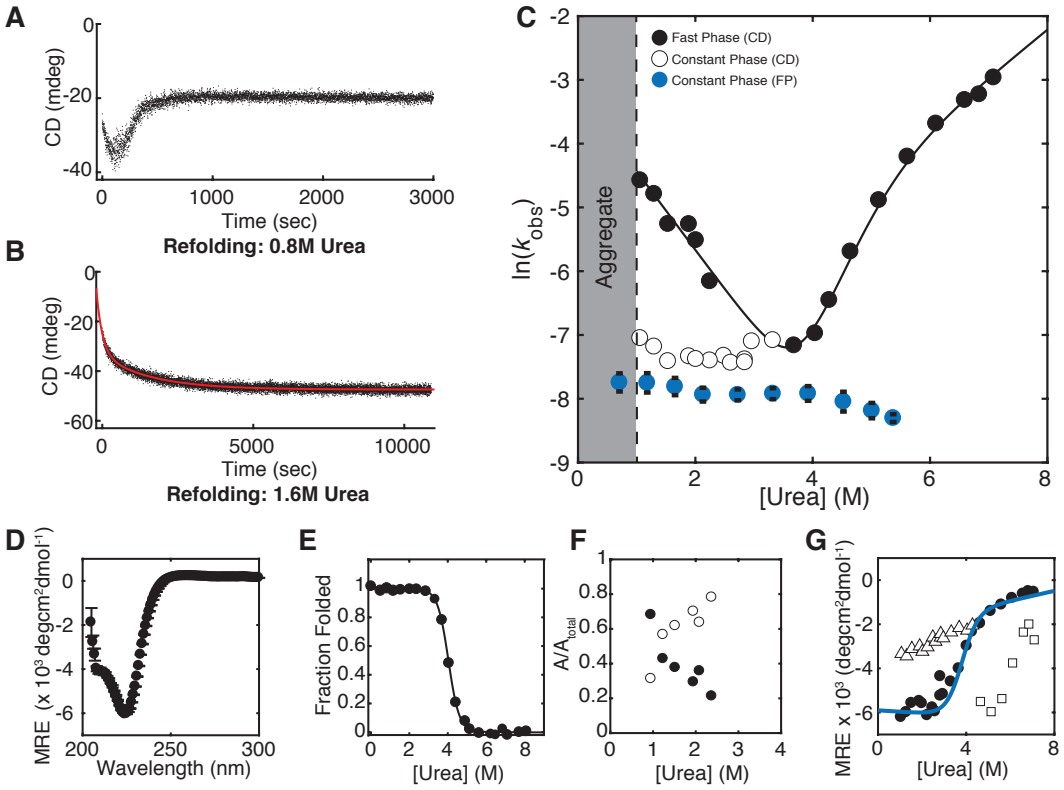


Figure 4.3 - Biophysical characterize of HaloTag's folding landscape. (A) Representative refolding trace at 0.8M urea, where HaloTag aggregates. **(B)** Representative refolding trace at urea concentrations above 1.0M urea. Red line is fit to a double exponential equation. **(C)** Chevron showing the dependence of HaloTag folding and unfolding rates as a function of urea concentration. HaloTag folding occurs in two phases as measured by CD, a fast phase (black circles) and a constant phase (white circles, black outline). Folding as measured by FP is shown in blue. Error bars represent the SD of three separate experiments. **(D)** CD spectrum of HaloTag at 0M urea. Error bars represent the SD of three separate experiments. **(E)** Equilbrium melt of HaloTag (fitting parameters are available in Table S1). **(F)** Normalized amplitudes of HaloTag constant phase (white circles, black outline) and urea-dependent fast folding phase (solid black circles). **(G)** Burst phase amplitudes for folding (white triangles with black outline) and unfolding (white squares with black outline) suggest the presence of an unfolding intermediate. Kinetic final amplitudes (black circles) overlay well with the fit of equilibrium data (blue line).

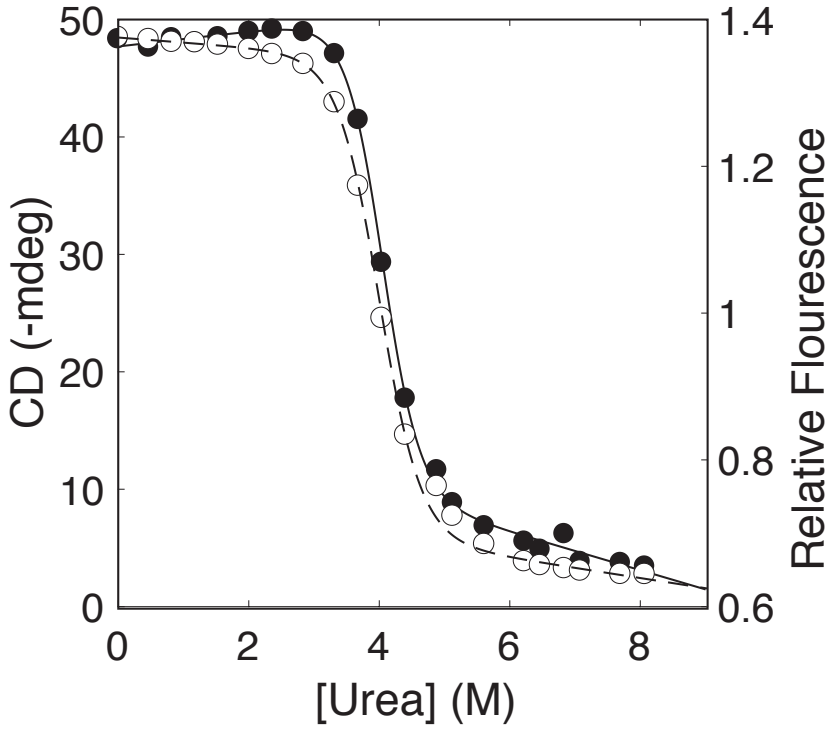
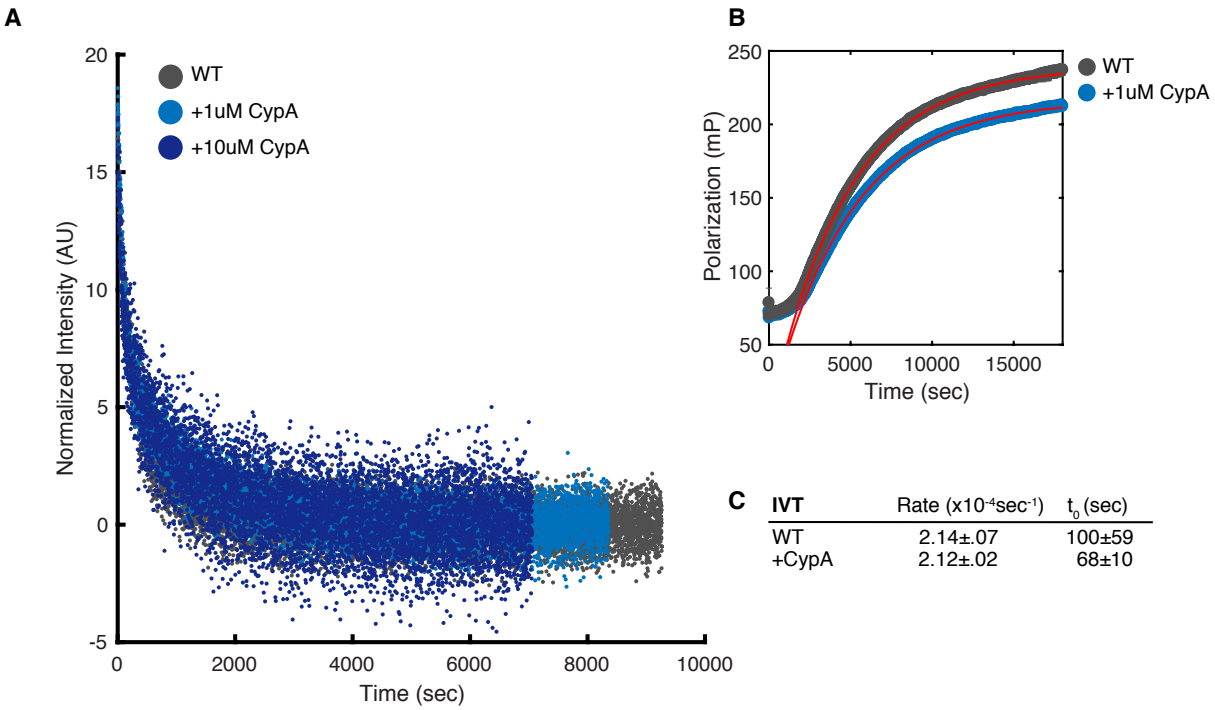


Figure 4.4 — Comparison of equilibrium denaturation as measured by CD (Black circles, black line) and fluorescence (White circles, dotted line).



C

IVT	Rate (x10^{-4}sec^{-1})	t_0 (sec)
WT	2.14±.07	100±59
+CypA	2.12±.02	68±10

Figure 4.5 – Addition of the peptidyl-proline isomerase Cyclophilin A (CypA) does not affect HaloTag *de novo* or refolding rates. (A) Refolding traces at increasing concentrations of CypA (grey- no CypA; light blue- +1uM CypA; dark blue- +10uM CypA). **(B)** IVT in the presence of CypA **(C)** Comparison of *de novo* folding rates as measured by IVT and FP in the presence or absence of CypA.

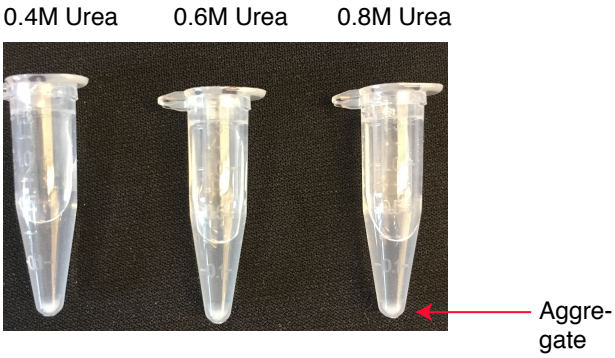


Figure 4.6 – HaloTag aggregates after refolding via dilution from 8.0M urea to the indicated final concentrations of urea.

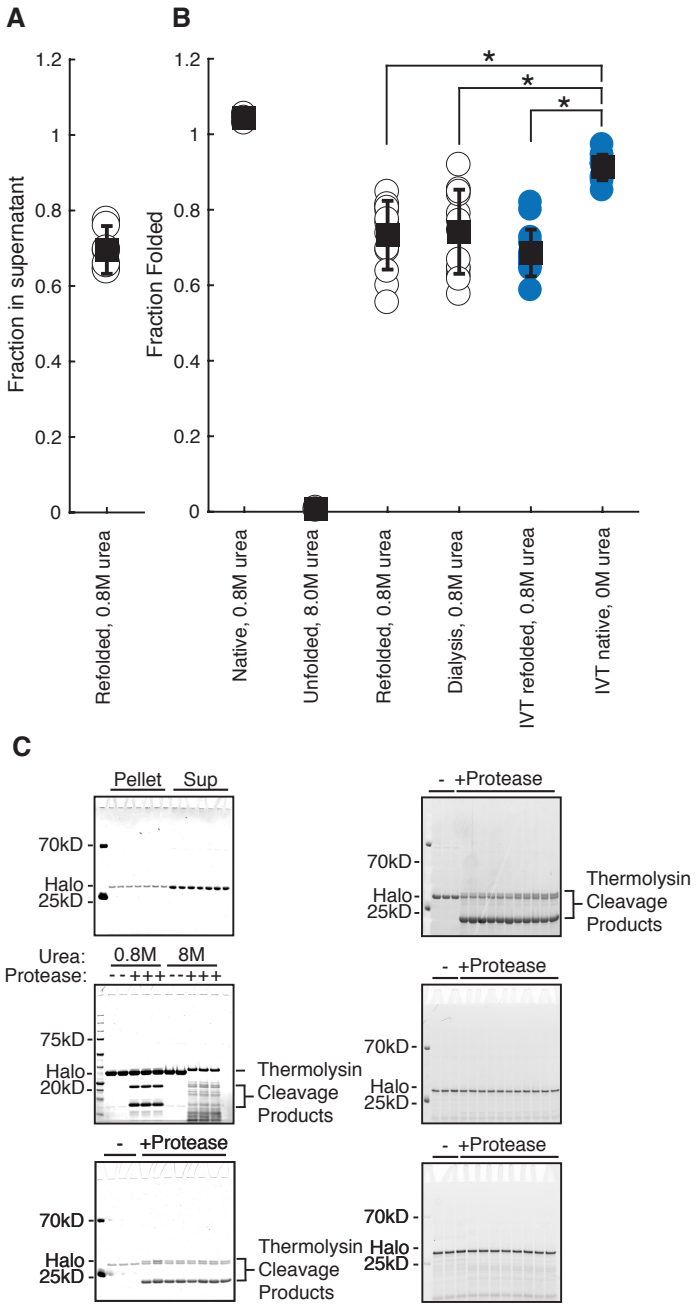


Figure 4.7 – HaloTag folding is more efficient during *in vitro translation* than after refolding (A) Fraction of total protein remaining in supernatant after refolding to 0.8M urea. **(B)** Fraction folded as measured by pulse proteolysis in conditions as indicated either after refolding, after *in vitro* translation or both. Blue circles are *in vitro* translated protein. **(C)** Representative gels for panels **(A)** and **(B)**. All error bars are the SD of at least 15 separate experiments except for HaloTag in 0.8 and 8.0M urea, which are the SD of three experiments. **'*'** represents a p-value of less than .01, using a student's unpaired t-test.

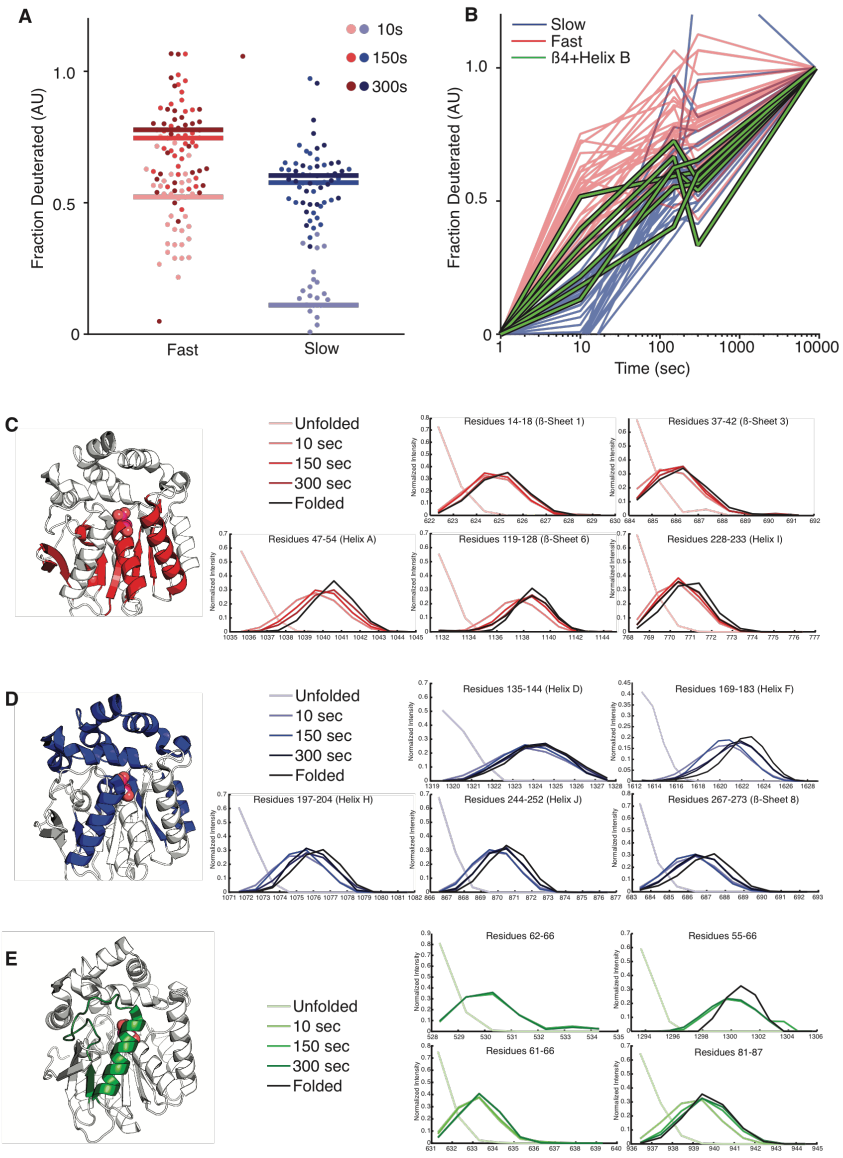


Figure 4.8 – Hydrogen Exchange/Mass Spectrometry reveals the structural basis for two distinct HaloTag folding phases. (A) Peptides were divided into fast or slow folding regions based on the fraction deuterated at the ten second time point. Mean fractions deuterated for each analyzed peptide (solid lines) are shown for three time points, while individual peptides are shown as filled circles. **(B)** Plot of fraction deuterated over time for each peptide used in this analysis. Fast in red, slow in blue and HelixB+ß-sheet 4 in green. **(C)** Structural representation of fast folding regions of HaloTag (red). Peptides for select regions of the protein are shown for each time point. Unfolded is in light red and folded is in black. Other time-points are colored as noted. **(D)** Structural representation of slow folding regions of HaloTag (red). Peptides for select regions of the protein are shown for each time point. Unfolded is in light blue and folded is in black. Other time-points are colored as noted. **(E)** Structural representation of Helix-B and ß-sheet 4 of HaloTag (green). Peptides for select regions of the protein are shown for each time point. Unfolded is in light green and folded is in black.

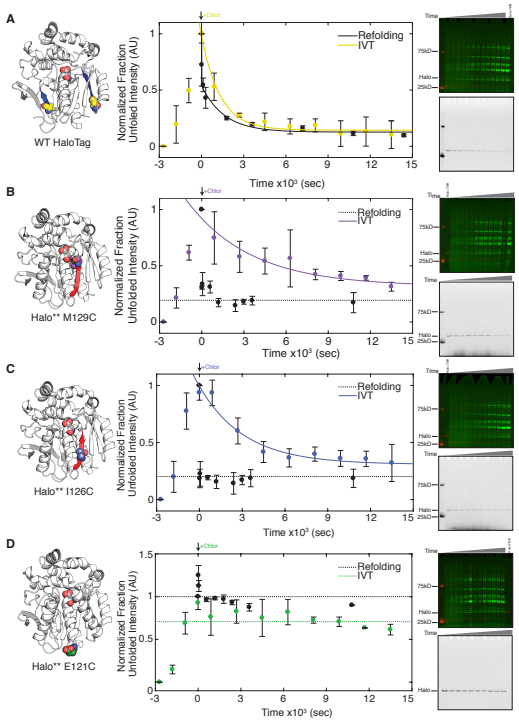
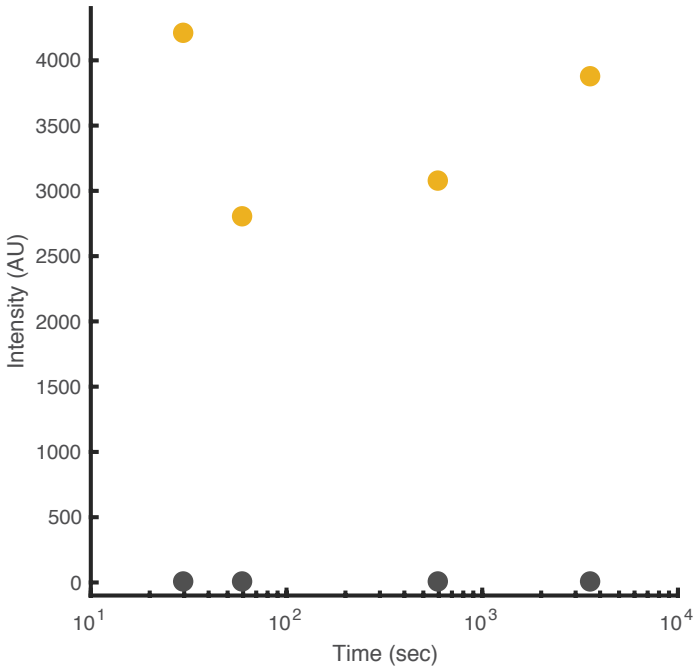


Figure 4.9 – Cysteine protection reveals that HaloTag cysteine accessibility is altered during translation. (A) HaloTag native cysteines (yellow spheres) are both located on slow folding regions of the protein (blue sheets). Asp106 is shown in white spheres as a reference. Cysteine protection was assessed during refolding (black line, black filled circles) or after the addition of chloramphenicol during IVT (yellow line, yellow spheres). **(B)** HaloTag** M129C (purple spheres) is located on a fast-folding region of the protein (red sheet). Asp106 is shown in white spheres as a reference. Cysteine protection was assessed during refolding (dotted line black filled circles) or after the addition of chloramphenicol during IVT (purple line, purple spheres). Note immediate protection of M129C during refolding but not during *in vitro* translation. **(C)** HaloTag** I126C (purple spheres) is located on a fast-folding region of the protein (red sheet). Asp106 is shown in white spheres as a reference. Cysteine protection was assessed during refolding (dotted line, black filled circles) or after the addition of chloramphenicol during IVT (purple line, purple spheres). Note immediate protection of M126C during refolding but not during *in vitro* translation. **(D)** HaloTag** E121C (Green spheres) is located on an exposed loop (green). Asp106 is shown in white spheres as a reference. Cysteine protection was assessed during refolding (dotted line, green filled circles) or after the addition of chloramphenicol during IVT (purple line, purple spheres). Since E121C is exposed, its accessibility does not change during IVT or refolding. Gels show cysteine protection during IVT (top, colored) and during refolding (bottom, grayscale). Labelled HaloTag is shown in red for reference during IVT experiments. All error bars are SD of three separate experiments. All fits are fit to a single exponential and represented as solid lines. Dotted lines are guides, not fits.

A



B

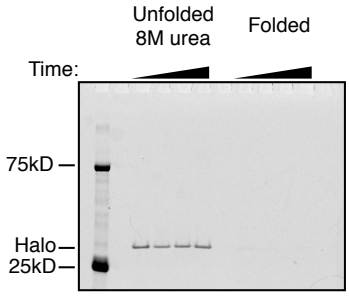


Figure 4.10 – HaloTag cysteines are completely protected in the folded state and are immediately labeled in the unfolded state (A) Flourescence intensity as a function of time of unfolded (yellow circles) and folded HaloTag (grey circles) as measured in **(B)**

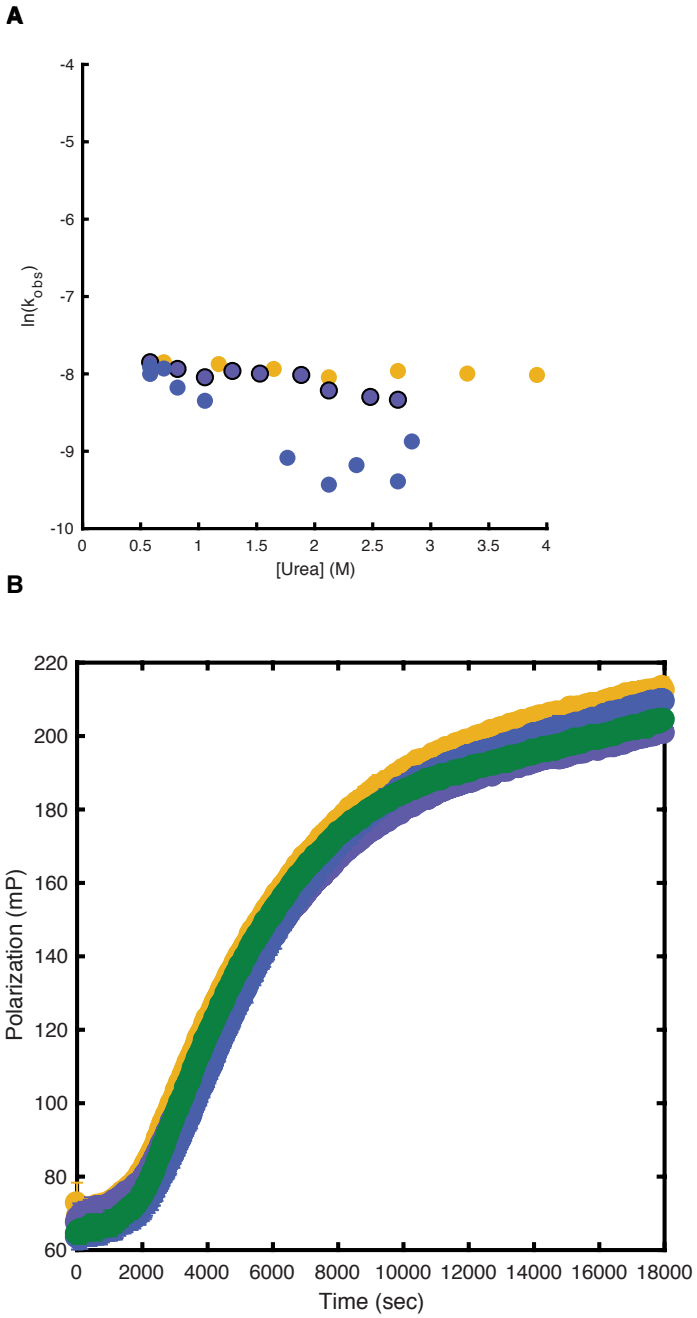


Figure 4.11 – HaloTag mutants fold at the same rate as WT as measured by FP during both refolding and *de novo* folding. (A)** Refolding rate as a function of urea as measured by FP. Yellow:WT; Dark purple with black outline: Halo** M129C; Light blue: Halo** I126C **(B)** Folding as measured by FP during *in vitro* translation. Yellow:WT; Dark purple with black outline: Halo** M129C; Light blue: Halo** I126C; Green: Halo** E121C

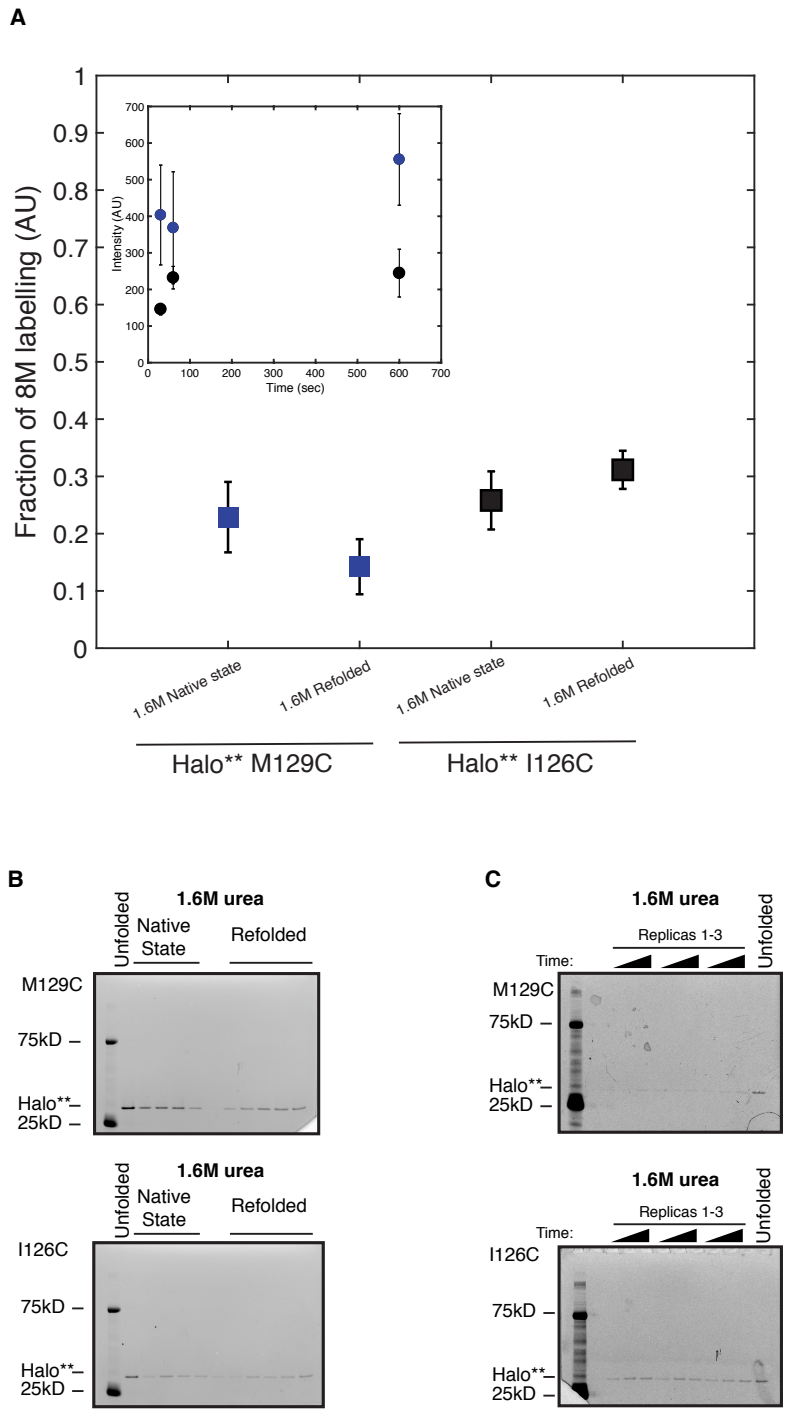


Figure 4.12 – HaloTag M129C and I126C native states are labelled to 20% of unfolded at 1.6M urea. (A)** Comparison of native state cysteine accessibility at 1.6M urea with cysteine accessibility after refolding and overnight equilibration for Halo** M129C (purple squares) and I126C (black squares). Inset is accessibility of the native state as a function of time for Halo** M129C (purple circles) and I126C (black circles). Gels for the native state experiment are shown in **(B)**. Gels for the time-dependent experiment are shown in **(C)**.

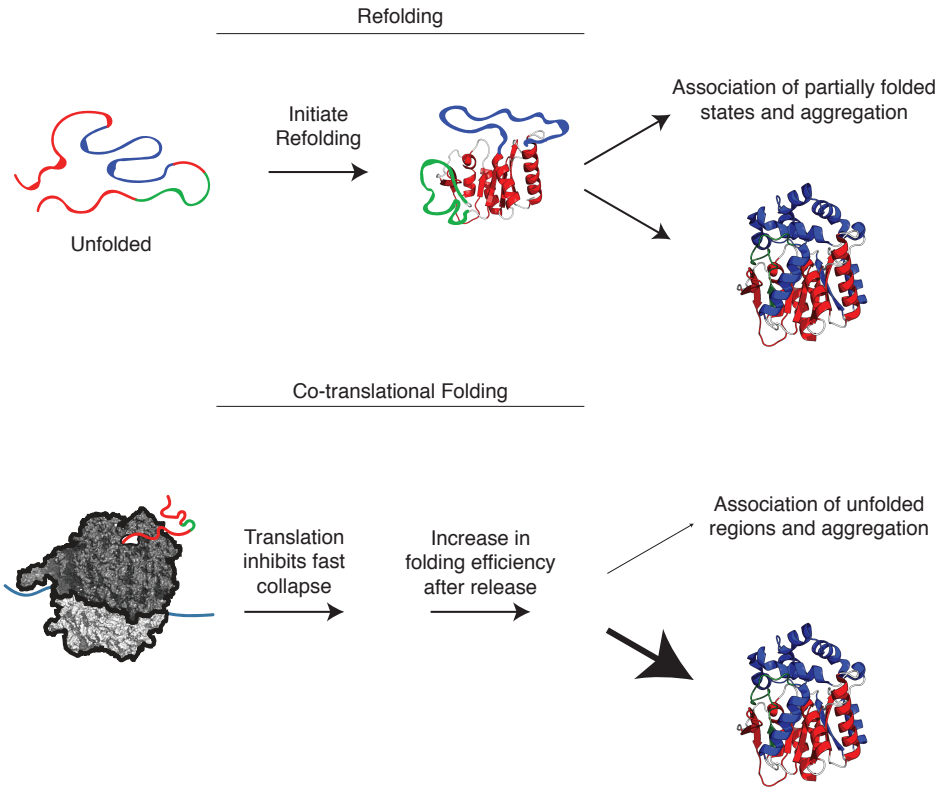


Figure 4.13 – Proposed model of HaloTag folding. During refolding (top), HaloTag has its entire sequence available and so its beta-sheet core forms quickly with the exception of sheet 4 and helix B. Since docking of ß4 and Helix B is slow, this leads to self-association of unfolded portions of HaloTag and aggregation. During *de novo* or co-translational folding (bottom), the beta-sheet core cannot form and thus there is no aggregation-prone structure formation leading to increased HaloTag folding efficiency.

4.15 Tables

Table 4.1 – Summary of kinetic and thermodynamic data of HaloTag and its mutants.

Thermodynamic Data		
	ΔG_{melt} ($kcal^{-1}mol$)	6.03±0.39
	m-value ($kcal\ mol^{-1}M^{-1}$)	1.57±0.11
Kinetic Data		
	$\Delta G_{kinetic}$ ($kcal^{-1}\ mol$)	5.24±2.0
	m-$value_{kinetic}$ ($kcal\ mol^{-1}M^{-1}$)	1.41±0.58
	$k_{f,H2O}$ (sec^{-1})	0.04±0.02
	m_f ($kcal\ mol^{-1}M^{-1}$)	1.46±0.71
	$k_{constant,\ H2O}$ (sec^{-1})	$6.6±0.71x10^{-4}$
	$m_{constant}$ ($kcal\ mol^{-1}M^{-1}$)	0.02±0.1
	$k_{FP,\ H2O}$ (sec-1)	$4.7±0.9x10^{-4}$
	m_{FP} ($kcal\ mol^{-1}M^{-1}$)	-0.1±0.04
	$k_{NI,\ H2O}$ (sec^{-1})	$8.47±20±10^{-6}$
	m_{NI} ($kcal\ mol^{-1}M^{-1}$)	-0.44±0.3
	$k_{IU,\ H2O}$ (sec^{-1})	$3.3±9.9x10^{-4}$
	m_{IU} ($kcal\ mol^{-1}M^{-1}$)	0.70±.15
Cysteine Accessibility		
	$k_{WT,refolding,\ slow}$ (sec^{-1})	$7.8±0.6±10^{-4}$
	$k_{WT,refolding,\ slow}$ (sec^{-1})	0.03±.018
	$k_{M129C,refolding}$ (sec^{-1})	<0.01
	$k_{I126C,refolding}$ (sec^{-1})	<0.01
	$k_{E121C,refolding}$ (sec^{-1})	NA
	$k_{WT,IVT}$ (sec^{-1})	$4.7±0.3x10^{-4}$
	$k_{M129C,IVT}$ (sec^{-1})	$3.2±0.2x10^{-4}$
	$k_{I126C,IVT}$ (sec^{-1})	$2.2±0.2x10^{-4}$
	$k_{E121C,IVT}$ (sec^{-1})	NA

Table 4.2 – Kinetic data obtained for HaloTag using fluorescence polarization (FP).

	Lag Time (sec)	Rate ($x10^{-4}$ sec^{-1})
Refolding (polarization)	NA	4.79±0.06
Folding (*in vitro* translation)	2009.3±72.7	2.67±0.07
Translation	275±57.2	5.78±0.70

Table 4.3 – Determination of HaloTag folding efficiency under different conditions

	Fraction Folded	Number of samples
Native	1.04±0.01	3
Unfolded	0.0006±0.003	3
0.8M refolded	0.73±0.10	15
0.8M dialysis	0.74±0.11	15
IVT refolded	0.69±0.06	15
IVT native	0.91±0.03	15

Chapter 5
Conclusions: New challenges and open questions in the study of nascent chain folding

5.1 Where do we go from here?

The idea that translation and folding are coupled at both a cellular-network level and single-protein level is not new[1,2]. Over the past thirty-or-so years, it has become clear that protein quality control in the cell is coordinated by a number of factors, and that those factor differ especially between eukaryotic and prokaryotic systems[3,4]. There are common themes and organizational centers, however, for dealing with proteostatic stress and the ribosome serves as a major protestatic coordination center in almost all organisms whose proteostasis response has been studied in detail[5,6]. Recently, there has been an increased interest in deciphering the players of both cellular proteostasis and protein (mis)folding in order to understand and treat a variety of diseases including Alzheimer's, ALS and aging [7–10]. Our knowledge of how protein chaperones and other actively regulated cellular machinery modulate their function in response to changing proteostatic conditions has grown significantly in recent years[11]. We now have a good idea which factors play a role in responding to specific types of proteostatic insult. We still do not have, however, a complete view of a protein's folding landscape in the cell. Why do changes in proteostasis lead to production of toxic products? What, at a physical level, makes a protein conformation toxic or not? Why is translation and the translational machinery (especially the ribosome) so important for managing proteostatic stress[6]? We cannot begin to answer these questions without understanding how translation and the ribosome alter the folding landscape of the emerging protein at a biophysical level. Until recently, these questions were answered for specific systems, using clever techniques that, unfortunately, were not always applicable to more than one protein[12–16].

Here, I have presented two new techniques for studying the interplay between the ribosome, translation and the emerging nascent chain and revealed some general properties of nascent chain folding. The first of these, combining pulse proteolysis with commercially available IVT and fluorescent labelling, provides a robust, general method for measuring both the stabilities and unfolding rates of ribosome nascent chains. The second, while dependent on one protein for its read-out, can be used to test the effect of almost every factor that plays a role in translation, such as the concentration of tRNA concentration and amino acids, mRNA structure, codon bias and ribosome composition.

5.2 New questions about how the ribosome modulates nascent chain energy landscapes

In chapter 2, we discovered that the ribosome exerts a destabilizing effect on the emerging chain and that this destabilization is dependent on the amino-acid distance between nascent chain and the peptidyl-transferase center (PTC). Then in chapter 3, we adapted this pulse proteolysis based technique to study the unfolding of a two-state

RNase H mutant, I53D, and were able to probe whether the destabilization is primarily a deceleration of folding or an acceleration of unfolding. For RNase H I53D, we found that it is acceleration of unfolding, rather than deceleration of folding that is more affected by the ribosome. One explanation for this has to do with the timescales of translation and folding. During translation, which occurs on the order of minutes, the nascent chain has time to undergo secondary structure formation and tertiary collapse. At the same time, the nascent chain physically cannot fold to its native state, as its full sequence is not synthesized yet. Furthermore, it is energetically favorable for hydrophobic residues to be buried. Thus, the nascent chain may undergo non-native collapse during synthesis and that there is the potential for the protein to become trapped in an off-pathway conformation. The destabilization we observe, and the fact that it is primarily due to an acceleration of unfolding, suggests that the ribosome has evolved to promote the resolution of partially folded nascent chain conformations. Supporting this hypothesis is the fact that Trigger Factor (TF), a prokaryotic ribosome associated chaperone, is known to "hold-and-unfold" proteins as they emerge from the ribosome, but only after about 60-90 residues have emerged[17–19]. The destabilization effect we observe is abrogated at the beginning of TF's range, suggesting a handoff between the ribosome's passive destabilization and TF's active unfoldase role. While this observation and model are intriguing, they do not provide a physical explanation for this effect.

Many groups have suggested that electrostatic interactions between the ribosome and the nascent chain are responsible for differences between nascent chain properties and properties of the free protein, but their physical basis still remains unknown[14,20,21]. Is there electrostatic repulsion of negatively charged residues? Our results demonstrating that DHFR is the most destabilized protein of the three we tested support this hypothesis **(Figure 3.4)**. It is also possible, however, that positively charged residues bind to the ribosome and thus favor an extended conformation, thus destabilizing the nascent chain. This has been observed using NMR for an IgG-like domain[20]. Fortunately, our method allows for the rapid determination of RNC stability, and thus has enabled experiments such as: (1) measuring the stability of RNCs in the presence of solutes with different Debye lengths (to mitigate or enhance electrostatic screening) (2) measure the stability of protein variants with vastly different isoelectric points (3) measuring the stability of proteins attached to inert microspheres with differentially charged surface-coatings **(Figure 5.1)**. Understanding the physical basis for changes in RNC biophysical properties has wide-ranging implications for protein design, folding simulations and understanding ribosome-associated protein quality control *in vivo*.

As stated above, a major underlying question in biology is whether protein folding trajectories are different *in vivo*, during translation, and *in vitro*. There are several cases where it is clear that the folding trajectory has been altered during translation. We continue to have a dearth of structural information for these folding trajectories: Which residues form structure when during translation? A powerful technique for understanding which residues are important in a protein's folding trajectory is the calculation of phi-values[22,23]. Using our new method, it is now possible to calculate phi-

values of ribosome nascent chains and compare them to those of proteins free in solution. By measuring the unfolding rates and stabilities of a two-state protein, it is possible to calculate the folding rate and thus, by extension, phi-values of its variants to determine residues involved in its folding transition state **(Figure 5.2)**. This would give us unprecedented insight into the mechanisms that govern folding on the ribosome. These experiments, however, are still only an approximation of the interplay between nascent chain folding and translation because they rely on stalled nascent chains rather than actively translating ribosomes.

5.3 Towards a high-resolution view of HaloTag folding during translation and refolding

Translation is a complex kinetic process that is dependent on large number of moving parts and de-convoluting nascent chain folding from the rest of the translational apparatus remains a challenge. HaloTag, which irreversibly binds ligands that can be functionalized with a variety of bio-orthogonal molecules, is a potent model system for studying complex processes that involve heterogeneous combinations of biomolecules. We have shown that HaloTag folding involves a urea-independent slow-folding phase that is necessary for ligand binding. Thus, ligand binding can be used as a measure of protein folding. Since HaloTag's slow folding phase is slower than translation, it is a natural system for studying co-translational folding. Suprisingly, HaloTag has several interesting folding properties, some of which seem to be modulated by translation (see Chapter 4). We characterized HaloTag folding using both CD and HX-MS and uncovered that: (1) HaloTag folds through at least one intermediate (2) HaloTag unfolding may also proceed through an intermediate (3) HaloTag's refolding efficiency is dependent on the final urea concentration of the refolding reaction (4) HaloTag folding during translation is more efficient than during refolding at the same urea concentrations (5) HaloTag's folding trajectory is different during translation than it is during refolding. HaloTag's folding trajectory during translation, however, is still structurally uncharacterized. Using both HXMS and cysteine accessibility experiments during translation, it should be possible to obtain a high-resolution picture of HaloTag's conformations as translation proceeds. This is also true of HaloTag's trajectory towards aggregate as it refolds in low urea conditions. Using mutagenesis informed by our HX-MS studies combined with stopped-flow folding kinetics studies, it should be possible to understand which regions of HaloTag are responsible for aggregation during refolding. Furthermore, similar types of studies could be done to reveal the structural elements responsible for HaloTag's slow kinetic folding phase **(Figure 5.3)**.

Finally, HaloTag is a great system for studying how the translational machinery itself modulates protein folding and opens a wide array of new experiments. What happens to folding if translation elongation instead of initiation is rate limiting? How does codon bias change HaloTag's folding trajectory? What about ribosome associated chaperones? HaloTag also is highly multiplexable and, considering its interesting folding properties, a great candidate for undertaking large screens to find determinants of protein stability, folding trajectories and aggregation. It is also possible to screen

libraries of HaloTag mutants in order to find variants that fold faster than translation and that eliminate HaloTag's constant folding phase. Furthermore, by generating a split HaloTag, it should be possible to apply these types of assays to a wide variety of proteins **(Figure 5.4)**.

5.4 Final Remarks

The holy-grail of protein folding is to predict protein structure, dynamics, and behavior simply from primary sequence. Here, we have introduced several new techniques for understanding quantitatively how the protein synthesis machinery might limit the conformational search of the emerging protein. In the future, I hope these techniques can be used to uncover general principles of nascent chain folding. Hopefully these discoveries can be incorporated into folding simulations to extend their timescale and accuracy. At the same time, the ribosome stands at a major quality control node in the cell and integrates a multitude of signals from intracellular signals in order to maintain cellular homeostasis. We have observed just a few ways in which the ribosome modulates nascent chain folding: (1) it destabilizes nascent chains (2) it changes both protein folding and unfolding rates and (3) it can drive the folding trajectory of a protein towards one with greater efficiency **(Figure 5.6)**. There are still questions remaining, however: How does the ribosome modulate protein stability? What is the mechanism by which the ribosome can change folding trajectories? Do partially folded intermediates have longer or shorter lifetimes during translation? I hope that this dissertation has made a small contribution to both expanding our understanding of ribosome modulated protein folding and also has made available new techniques for quantitatively understanding general principles of protein folding.

5.5 References

1. Purvis IJ, Bettany AJE, Santiago TC, et al. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol*. 1987;193(2):413-417. doi:10.1016/0022-2836(87)90230-0.
2. Baldwin RL. Intermediates in protein folding reactions and the mechanism of protein folding. *Annu Rev Biochem*. 1975;44(1):453-475. doi:10.1146/annurev.bi.44.070175.002321.
3. Brandvold KR, Morimoto RI. The Chemical Biology of Molecular Chaperones--Implications for Modulation of Proteostasis. *J Mol Biol*. 2015;427(18):2931-2947. doi:10.1016/j.jmb.2015.05.010.
4. Labbadia J, Morimoto RI. The Biology of Proteostasis in Aging and Disease. *Annu Rev Biochem*. 2015;(March):1-30. doi:10.1146/annurev-biochem-060614-033955.
5. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. *Nature*. 2011;475(7356):324-332. doi:10.1038/nature10317.
6. Pechmann S, Willmund F, Frydman J. The ribosome as a hub for protein quality control. *Mol Cell*. 2013;49(3):411-421. doi:10.1016/j.molcel.2013.01.020.
7. Balch WE, Morimoto RI, Dillin A, Kelly JW. Adapting proteostasis for disease

intervention. *Science*. 2008;319(5865):916-919. doi:10.1126/science.1141448.
8. Roth DM, Balch WE. Modeling general proteostasis: proteome balance in health and disease. *Curr Opin Cell Biol*. 2011;23(2):126-134. doi:10.1016/j.ceb.2010.11.001.
9. Stefani M, Dobson CM. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J Mol Med*. 2003;81(11):678-699. doi:10.1007/s00109-003-0464-5.
10. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426(6968):884-890. doi:10.1038/nature02261.
11. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591.
12. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858-861. doi:10.1021/ja411302m.
13. Clark PL, King J. A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem*. 2001;276(27):25411-25420. doi:10.1074/jbc.M008490200.
14. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C. The ribosome modulates nascent protein folding. *Science*. 2011;334(6063):1723-1727. doi:10.1126/science.1209740.
15. Holtkamp W, Kokic G, Jäger M, Mittelstaet J, Komar AA, Rodnina M V. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104-1107. doi:10.1126/science.aad0344.
16. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. *Curr Opin Struct Biol*. 2010;20(1):33-45. doi:10.1016/j.sbi.2010.01.005.
17. Oh E, Becker AH, Sandikci A, et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*. 2011;147(6):1295-1308. doi:10.1016/j.cell.2011.10.044.
18. Hoffmann A, Becker AH, Zachmann-Brand B, Deuerling E, Bukau B, Kramer G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol Cell*. 2012;48(1):63-74. doi:10.1016/j.molcel.2012.07.018.
19. Nilsson OB, Müller-Lucks A, Kramer G, Bukau B, von Heijne G. Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein. *J Mol Biol*. 2016;428(6):1356-1364. doi:10.1016/j.jmb.2016.02.014.
20. Cabrita LD, Cassaignau AME, Launay HMM, et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol*. 2016;(February). doi:10.1038/nsmb.3182.
21. Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, Cavagnero S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem Biol*. 2013;8(6):1195-1204. doi:10.1021/cb400030n.
22. de los Rios M a, Muralidhara BK, Wildes D, et al. On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci*. 2006;15(3):553-563. doi:10.1110/ps.051870506.

23. Matouschek A, Kellis JT, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. *Nature*. 1989;340(6229):122-126. doi:10.1038/340122a0.

5.6 Figures

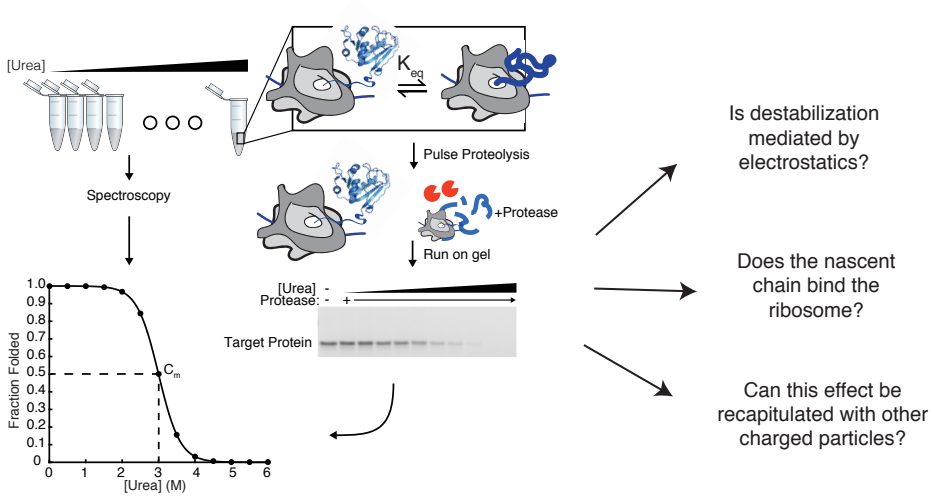


Figure 5.1 – New questions that are answerable using pulse proteolysis to probe RNC stability.

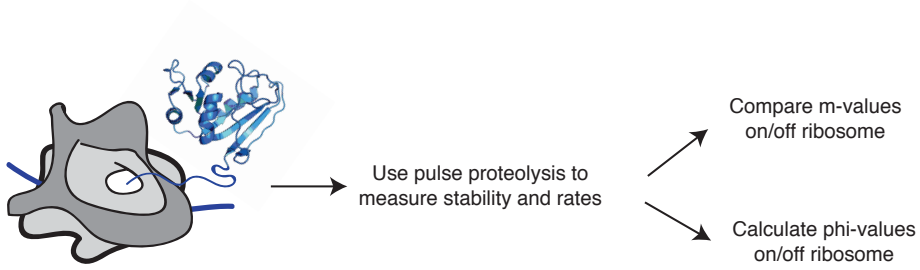


Figure 5.2 – Determining RNC unfolding kinetics by pulse proteolysis will comparison of m-values and phi-values on/off the ribosome, shedding new light on how the ribosome changes protein folding trajectories.

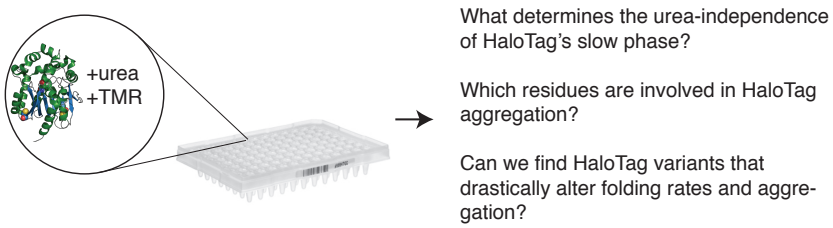


Figure 5.3 – High-throughput measurement of HaloTag's protein folding landscape has the potential to reveal further details of its folding and aggregation.

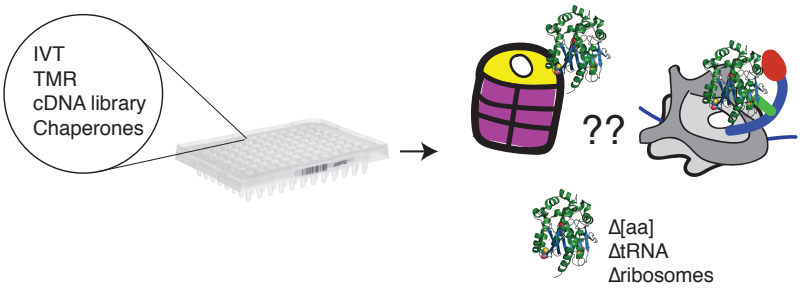


Figure 5.4 – HaloTag can be used in a high-throughput manner to determine which translation parameters affect protein folding as well as how chaperones such as trigger factor (blue and red) and GroEL/ES (purple and yellow) change HaloTag's folding landscape.